5   NOVEL GENES ENCODING PROTEINS HAVING

DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER USES

Cross Reference to Related Applications

This application is a continuation-in-part of co-pending United States

10   Patent Application number 09/342,364, filed June 29, 1999.

Background of the Invention

The molecular bases underlying many human and animal physiological

states (e.g., diseased and homeostatic states of various tissues) remain unknown.

15   Nonetheless, it is well understood that these states result from interactions among the

proteins and nucleic acids present in the cells of the relevant tissues. In the past, the

complexity of biological systems overwhelmed the ability of practitioners to

understand the molecular interactions giving rise to normal and abnormal physiological

states. More recently, though, the techniques of molecular biology, transgenic and null

20   mutant animal production, computational biology, pharmacogenomics, and the like

have enabled practitioners to discern the role and importance of individual genes and

proteins in particular physiological states.

Knowledge of the sequences and other properties of genes (particularly

including the portions of genes encoding proteins) and the proteins encoded thereby

25   enables the practitioner to design and screen agents which will affect, prospectively or

retrospectively, the physiological state of an animal tissue in a favorable way. Such

knowledge also enables the practitioner, by detecting the levels of gene expression and

protein production, to diagnose the current physiological state of a tissue or animal and

to predict such physiological states in the future. This knowledge furthermore enables

30   the practitioner to identify and design molecules which bind with the polynucleotides

and proteins, *in vitro, in vivo,* or both.

The present invention provides sequence information for polynucleotides derived from human genes and for proteins encoded thereby, and thus enables the practitioner to assess, predict, and affect the physiological state of various human tissues.

5   Summary of the Invention

The present invention is based, at least in part, on the discovery of a variety of human cDNA molecules which encode proteins which are herein designated INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, and TANGO 332. These seven proteins, fragments thereof, derivatives

10   thereof, and variants thereof are collectively referred to herein as the polypeptides of the invention or the proteins of the invention. Nucleic acid molecules encoding polypeptides of the invention are collectively referred to as nucleic acids of the invention.

The nucleic acids and polypeptides of the present invention are useful as

15   modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, the present invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present invention also provides nucleic acid molecules which are suitable as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the

20   invention.

The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, the present invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present

25   invention also provides nucleic acid molecules which are suitable as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention.

The invention also features nucleic acid molecules which are at least 40% (or 50%, 60%, 70%, 80%, 90%, 95%, or 98%) identical to the nucleotide

sequence of any of SEQ ID NOs: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or the nucleotide sequence of a cDNA clone deposited with ATCC® as one of Accession numbers PTA-147, PTA-150, 207230, and PTA-151 ("a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151"), or a complement

5   thereof.

The invention features nucleic acid molecules which include a fragment of at least 15 (25, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 4500, or 4928) consecutive nucleotide residues of any of SEQ ID NOs: 1, 2, 9,

10   10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least

15   50% (or 60%, 70%, 80%, 90%, 95%, or 98%) identical to the amino acid sequence of any of SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, or the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, or a complement thereof.

In preferred embodiments, the nucleic acid molecules have the

20   nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151.

Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of any of SEQ ID NOs: 3-8,

25   11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, or the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, the fragment including at least 8 (10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, or 200) consecutive amino acids of any of SEQ ID NOs: 3-8, 11-32, 35-

- 3 -

37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, or the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151.

The invention includes nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® PTA-147, PTA-150, 207230, and PTA-151, wherein the nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule having a nucleic acid sequence encoding any of SEQ ID NOs: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, or a complement thereof.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 50%, preferably 60%, 75%, 90%, 95%, or 98% identical to the amino acid sequence of any of SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 40%, preferably 50%, 75%, 85%, or 95% identical the nucleic acid sequence encoding any of SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule consisting of the nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92.

Also within the invention are polypeptides which are naturally occurring allelic variants of a polypeptide that includes the amino acid sequence of any of SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, or the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes under stringent conditions to a nucleic acid molecule having the

- 4 -

nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, or a complement thereof. In other embodiments, the nucleic acid molecules are at least 15 (25, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 4500, or 4928) nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, or a complement thereof. In some embodiments, the isolated nucleic acid molecules encode a cytoplasmic, transmembrane, extracellular, or other domain of a polypeptide of the invention. In other embodiments, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a nucleic acid of the invention.

Another aspect of the invention provides vectors, e.g., recombinant expression vectors, comprising a nucleic acid molecule of the invention. In another embodiment, the invention provides isolated host cells, e.g., mammalian and non-mammalian cells, containing such a vector or a nucleic acid of the invention. The invention also provides methods for producing a polypeptide of the invention by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector encoding a polypeptide of the invention such that the polypeptide of the invention is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides of the invention. Preferred proteins and polypeptides possess at least one biological activity possessed by the corresponding naturally-occurring human polypeptide. An activity, a biological activity, and a functional activity of a

- 5 -

polypeptide of the invention refers to an activity exerted by a protein or polypeptide of the invention on a responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein or an indirect activity, such as a

5 cellular processes mediated by interaction of the protein with a second protein.

By way of example, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof exhibit the ability to affect growth, proliferation, survival, differentiation, and activity of human pancreas, skeletal muscle, heart, brain, placenta, lung, liver, and kidney cells. INTERCEPT 217 modulates cellular binding to one or

10 more mediators, modulates activity and release of one or more pancreatically secreted digestive enzymes, and protects tissue from endogenous digestive enzymes. Thus, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prevent, diagnose, or treat disorders relating to aberrant endogenous digestive enzyme activity, inappropriate interaction (or non-interaction) of cells with mediators,

15 inappropriate cellular development and proliferation, inappropriate inflammation, and inappropriate immune responses. Exemplary disorders for which INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof are useful include immune disorders (e.g., insufficient immune responses and auto-immune disorders), infectious diseases, auto-immune disorders, pancreatic disorders (e.g., pancreatitis and pancreatic

20 carcinoma), disorders related to mal-expression of growth factors, cancers, inflammatory disorders, acute and chronic traumas, and the like.

Further by way of example, INTERCEPT 297 polypeptides, nucleic acids, and modulators thereof exhibit the ability to affect growth, proliferation, survival, differentiation, and activity of human fetal cells and spleen cells and of (e.g.,

25 bacterial or fungal) cells and viruses which infect humans. Furthermore, INTERCEPT 297 modulates organization, structure, and function of biological membranes. Thus, INTERCEPT 297 polypeptides, nucleic acids, and modulators thereof can be used to affect development and persistence of atherogenesis and arteriosclerosis, for example, or to modulate transmembrane transport processes such as ion transport across neuronal

and muscle cell membranes (e.g., ion transport relating to nerve impulse conduction and muscle contraction). INTERCEPT 297 polypeptides, nucleic acids, and modulators thereof can be used to prevent, diagnose, or treat transmembrane transport disorders such as cystic fibrosis, pain, seizure, epilepsy, mental disorders, and the like. Other exemplary disorders for which INTERCEPT 297 polypeptides, nucleic acids, and modulators thereof are useful include disorders involving generation and persistence of an immune response to bacterial, fungal, and viral infections.

Still further by way of example, TANGO 276 polypeptides, nucleic acids, and modulators thereof modulate growth, proliferation, survival, differentiation, and activity of human heart, placenta, brain, lung, liver, skin, kidney, pancreas, spleen, and fetal tissues. TANGO 276 guides neuronal growth and development and modulates growth, homeostasis, and regeneration of other epithelial tissues. TANGO 276 is a secreted protein which mediates cellular interaction with cells, molecules, and structures (e.g., extracellular matrix) in the extracellular environment. TANGO 276 is therefore involved in growth, organization, migration, and adhesion of tissues and the cells which constitute those tissues. Furthermore, TANGO 276 modulates growth, proliferation, survival, differentiation, and activity of neuronal cells and immune system cells. Thus, TANGO 276 polypeptides, nucleic acids, and modulators thereof can be used, for example, to prevent, diagnose, or treat disorders characterized by aberrant organization or development of a tissue or organ, for modulating migration and adhesion of cells (e.g., in disorders such as cancer metastasis, autoimmune disorders, and graft-versus-host disease or in normal or aberrant processes involving angiogenesis, such as tumor growth and persistence), for guiding neural axon development and regeneration, for modulating differentiation of cells of the immune system (e.g., to treat bacterial, fungal, or viral infection or to prevent, diagnose, or treat autoimmune disorders), for modulating cytokine production by cells of the immune system (e.g., to prevent, detect, or treat inflammation and pain), for modulating reactivity of cells of the immune system toward cytokines, for modulating initiation

- 7 -

and persistence of an inflammatory response, and for modulating proliferation of epithelial cells.

Yet further by way of example, TANGO 292 polypeptides, nucleic acids, and modulators thereof modulate growth, proliferation, survival, differentiation, and activity of human keratinocytes, including embryonic keratinocytes. TANGO 292, a transmembrane protein, is also involved in binding and uptake of calcium and other metal ions, and in responses of cells which express it to the presence and uptake of such ions. TANGO 292 polypeptides, nucleic acids, and modulators thereof can therefore be used to prevent, diagnose, and treat disorders involving one or more physiological activities mediated by TANGO 292 protein. These activities include, for example, bone uptake, maintenance, and deposition, formation, maintenance, and repair of cartilage and skin, formation and maintenance of extracellular matrices, movement of cells through extracellular matrices, coagulation and dissolution of blood components, and deposition of materials in and on arterial walls. TANGO 292 is also related to a variety of disorders which involve these activities. Such disorders include, for example, various bone-related disorders such as osteoporosis, skeletal development disorders, bone fragility, traumatic bone injuries, rickets, osteomalacia, Paget's disease, and other bone disorders, osteoarthritis, rheumatoid arthritis, ankylosing spondylitis, and other disorders of the joints and cartilage, skin disorders such as psoriasis, eczema, scleroderma, and skin tumors (e.g., keratoses, squamous cell carcinomas, malignant melanomas, and Kaposi's sarcomas), iron deficiency anemia, hemophilia, inappropriate blood coagulation, stroke, arteriosclerosis, atherosclerosis, aneurysm, and other disorders related to blood and blood vessels, metastasis and other disorders related to inappropriate movement of cells through extracellular matrices, and the like. TANGO 292 polypeptides, nucleic acids, and modulators thereof can thus be used to prevent, diagnose, and treat one or more of these disorders. TANGO 292 is also involved in skin disorders such as psoriasis, eczema, scleroderma, skin tumors (e.g., keratoses, squamous cell carcinomas, malignant melanomas, and Kaposi's sarcomas), in placental disorders such as placenta previa and *abruptio placentae*, in liver disorders such as

- 8 -

cirrhosis of the liver, liver fibrosis, hepatitis, and hepatic cancers, in kidney disorders such as urolithiasis, glomerulonephritis, nephrosis, renal cell carcinomas, and renal failure (both acute and chronic), in lung disorders such as cystic fibrosis, chronic obstructive pulmonary diseases (e.g., emphysema, bronchitis, and bronchiectasis), lung

5   cancers, and asthma, in pancreatic disorders such as diabetes, pancreatitis, pancreatic cancers, and pancreatic insufficiency, in cardiac disorders such as coronary artery disease (and other ischemic heart diseases), arrhythmia, congestive heart failure, endocarditis, and pericarditis, and the like. Thus, TANGO 292 polypeptides, nucleic acids, and modulators thereof can thus be used to prevent, diagnose, and treat one or

10   more of these disorders.

As an additional example, TANGO 325 polypeptides, nucleic acids, and modulators thereof modulate growth, proliferation, survival, differentiation, and activity of human tissues such as vascular endothelium, including aortic endothelium, other heart tissues, placenta, liver, kidney, and pancreas tissues. Thus, TANGO 325

15   polypeptides, nucleic acids, and modulators thereof can therefore be used to prevent, diagnose, and treat disorders involving one or more physiological activities mediated by TANGO 325 protein in tissues in which it is expressed. Such activities include, for example, modulation of cardiac contractility and vasomotor tone, modulation of leukocyte extravasation, sensing physiological signals by the endocrine system,

20   modulating growth, development, maintenance, and regeneration of neurons, and the like. Disorders related to these activities include, by way of example and not limitation, cardiovascular disorders such as arteriosclerosis, atherosclerosis, coronary artery disease (and other ischemic heart diseases), angina, myocardial infarction, restenotic disorders, hypertension, Buerger's disease, aneurysm, stroke, arrythmia,

25   congestive heart failure, endocarditis, and pericarditis, placental disorders such as placenta previa and *abruptio placentae*, liver disorders such as cirrhosis of the liver, liver fibrosis, hepatitis, and hepatic cancers, kidney disorders such as urolithiasis, glomerulonephritis, nephrosis, renal cell carcinomas, and renal failure (both acute and chronic), pancreatic disorders such as diabetes, pancreatitis, pancreatic cancers, and

pancreatic insufficiency, neurological system disorders, immune and auto-immune disorders, hyperthyroidism, hypothyroidism, diabetes, goiter, growth and developmental disorders, and the like.

Further by way of example, TANGO 331 polypeptides, nucleic acids, and modulators thereof modulate growth, proliferation, survival, differentiation, and activity of human fetal, lung, spleen, and thymus cells and tissues. As described herein, TANGO 331 is involved in physiological activities such as maintenance of epithelia, carcinogenesis, modulation and storage of protein factors and metals, lactation, and infant nutrition. TANGO 331 also modulates cellular binding and uptake of cytokines, growth factors, and metal ions. Thus, TANGO 331 polypeptides, nucleic acids, and modulators thereof can be used to prevent, diagnose, and treat disorders such as breast cancer, insufficient lactation, infant nutritional and growth disorders, malnutrition and mineral deficiency disorders, hemochromatosis, inappropriate calcification of body tissues, bone disorders such as osteoporosis, autoimmune disorders, insufficient or inappropriate host responses to infection, acquired immune deficiency syndrome, and the like.

As another example, TANGO 332 polypeptides, nucleic acids, and modulators thereof modulate growth, proliferation, survival, differentiation, and activity of human brain and other tissues. As described herein, TANGO 332 is involved in modulating establishment and maintenance of neural connections, cell-to-cell adhesion, tissue and extracellular matrix invasivity, and the like. Thus, TANGO 332 polypeptides, nucleic acids, and modulators thereof can be used to prevent, diagnose, and treat disorders such as brain cancers (e.g., gliomas, astrocytomas, medulloblastomas, ependymomas, Schwannomas, pituitary adenomas, teratomas, and the like), disorders of neural connection establishment or maintenance, impaired cognitive function, dementia, senility, Alzheimer's disease, mental retardation, inflammation, immune and autoimmune responses, and the like.

In one embodiment, a polypeptide of the invention has an amino acid sequence sufficiently identical to an identified domain of a polypeptide of the

invention. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 65% identity, preferably 75% identity, more preferably 85%, 95%, or 98% identity are defined herein as sufficiently identical.

In one embodiment, the isolated polypeptide of the invention lacks both a transmembrane and a cytoplasmic domain. In another embodiment, the polypeptide lacks both a transmembrane domain and a cytoplasmic domain and is soluble under physiological conditions.

The polypeptides of the present invention, or biologically active portions thereof, can be operably linked to a heterologous amino acid sequence to form fusion proteins. The invention further features antibody substances that specifically bind a polypeptide of the invention such as monoclonal or polyclonal antibodies, antibody fragments, single-chain antibodies, and the like. In addition, the polypeptides of the invention or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers. These antibody substances can be made, for example, by providing the polypeptide of the invention to an immunocompetent vertebrate and thereafter harvesting blood or serum from the vertebrate.

In another aspect, the present invention provides methods for detecting the presence of the activity or expression of a polypeptide of the invention in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of activity such that the presence of activity is detected in the biological sample.

In another aspect, the invention provides methods for modulating activity of a polypeptide of the invention comprising contacting a cell with an agent

- 11 -

that modulates (inhibits or enhances) the activity or expression of a polypeptide of the invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention.

In another embodiment, the agent modulates expression of a polypeptide of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense with respect to the coding strand of an mRNA encoding a polypeptide of the invention.

The present invention also provides methods to treat a subject having a disorder characterized by aberrant activity of a polypeptide of the invention or aberrant expression of a nucleic acid of the invention by administering an agent which is a modulator of the activity of a polypeptide of the invention or a modulator of the expression of a nucleic acid of the invention to the subject. In one embodiment, the modulator is a protein of the invention. In another embodiment, the modulator is a nucleic acid of the invention. In other embodiments, the modulator is a peptide, peptidomimetic, or other small molecule.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a polypeptide of the invention, (ii) mis-regulation of a gene encoding a polypeptide of the invention, and (iii) aberrant post-translational modification of a polypeptide of the invention wherein a wild-type form of the gene encodes a polypeptide having the activity of the polypeptide of the invention.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In general, such methods entail measuring a biological activity of the polypeptide in the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by measuring the expression of the polypeptide or nucleic acid in the presence and absence of the compound.

In yet a further aspect, the invention provides substantially purified antibodies or fragments thereof, including non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide having an amino acid sequence comprising a sequence selected from the group consisting of

(i) SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98;

(ii) the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151;

(iii) a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98;

(iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and

(v) an amino acid sequence which is encoded by a nucleic acid molecule, the complement of which hybridizes with a nucleic acid molecule having the sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or with a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, under conditions of hybridization of 6× SSC (standard saline citrate buffer) at 45°C and washing in 0.2× SSC, 0.1% SDS at 65°C.

In various embodiments, the substantially purified antibodies of the invention, or fragments thereof, can be human, non-human, chimeric and/or humanized antibodies.

- 13 -

In another aspect, the invention provides non-human antibodies or fragments thereof, which antibodies or fragments specifically bind with a polypeptide having an amino acid sequence comprising a sequence selected from the group consisting of

5   (i) SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98;

(ii) the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151;

(iii) a fragment of at least 15 amino acid residues of the amino acid sequence of

10   SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98;

(iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and

15   a gap penalty of 4; and

(v) an amino acid sequence which is encoded by a nucleic acid molecule, the complement of which hybridizes with a nucleic acid molecule having the sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or with a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, under

20   conditions of hybridization of 6× SSC (standard saline citrate buffer) at 45°C and washing in 0.2× SSC, 0.1% SDS at 65°C.

Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the

25   invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide having an amino acid sequence comprising a sequence selected from the group consisting of

- 14 -

(i) SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98;

(ii) the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151;

(iii) a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98;

(iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and

(v) an amino acid sequence which is encoded by a nucleic acid molecule, the complement of which hybridizes with a nucleic acid molecule having the sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or with a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, under conditions of hybridization of 6× SSC (standard saline citrate buffer) at 45°C and washing in 0.2× SSC, 0.1% SDS at 65°C.

The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

In a particularly preferred embodiment, the antibody substance of the invention specifically binds with an extracellular domain of one of INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, and TANGO 332. Preferably, the extracellular domain with which the antibody substance binds has an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 14-18, 37, 43, 51, 58, 63, 83, or 93.

Any of the antibodies of the invention can be conjugated with a therapeutic moiety or with a detectable substance. Non-limiting examples of detectable substances that can be conjugated with the antibodies of the invention include an

- 15 -

enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 comprises Figures 1A through 1M. The nucleotide sequence (SEQ ID NO: 1) of a cDNA encoding the human INTERCEPT 217 protein described herein is listed in Figures 1A through 1E. The open reading frame (ORF; residues 215 to 1579; SEQ ID NO: 2) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 3) of human INTERCEPT 217 is listed. Figure 1F is a hydrophilicity plot of human INTERCEPT 217 protein, in which the locations of cysteine residues ("Cys") and potential N-glycosylation sites ("Ngly") are indicated by vertical bars and the predicted extracellular ("out"), intracellular ("ins"), or transmembrane ("TM") locations of the protein backbone is indicated by a horizontal bar. An alignment of the amino acid sequences of human INTERCEPT 217 protein ("H"; SEQ ID NO: 3) and porcine ribonuclease inhibitor protein ("P"; SwissProt Accession number P10775; SEQ ID NO: 64) is shown in Figures 1G and 1H, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ".". These alignments were made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4). The nucleotide sequence (SEQ ID NO: 92) of an ORF encoding the murine INTERCEPT 217 protein described herein is listed in Figures 1I

- 16 -

through 1K. The ORF is indicated by nucleotide triplets, beneath which the amino acid sequence (SEQ ID NO: 93) of murine INTERCEPT 217 is listed. Figure 1L is a hydrophilicity plot of murine INTERCEPT 217 protein, in which the locations of cysteine residues ("Cys") and potential N-glycosylation sites ("Ngly") are indicated by

5   vertical bars and the predicted extracellular ("out"), intracellular ("ins"), or transmembrane ("TM") locations of the protein backbone is indicated by a horizontal bar. An alignment of the amino acid sequences of human INTERCEPT 217 protein ("H"; SEQ ID NO: 3) and murine INTERCEPT 217 protein ("M"; SEQ ID NO: 93) is shown in Figure 1M, wherein identical amino acid residues are indicated by " | " and

10   similar amino acid residues are indicated by ".". These alignments were made using the BESTFIT software (BLOSUM62 scoring matrix, gap opening pentaly = 12, frameshift gap penalty = 5, gap extension penalty = 4).

        Figure 2 comprises Figures 2A through 2D. The nucleotide sequence (SEQ ID NO: 9) of a cDNA encoding the human INTERCEPT 297 protein described

15   herein is listed in Figures 2A, 2B, and 2C. The open reading frame (ORF; residues 40 to 1152; SEQ ID NO: 10) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 11) of human INTERCEPT 297 is listed. Figure 2D is a hydrophilicity plot of human INTERCEPT 297 protein.

        Figure 3 comprises Figures 3A through 3R. The nucleotide sequence

20   (SEQ ID NO: 33) of a cDNA encoding the human TANGO 276 protein described herein is listed in Figures 3A to 3D. The ORF (residues 58 to 786; SEQ ID NO: 34) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 35) of human TANGO 276 is listed. Figure 3E is a hydrophilicity plot of TANGO 276 protein. An alignment of the amino acid sequences of human TANGO

25   276 protein ("H"; SEQ ID NO: 35) and murine protein M-Sema-F ("M"; SEQ ID NO: 65) is shown in Figures 3F to 3H. In Figures 3I through 3R, an alignment of the nucleotide sequences of the cDNA encoding human TANGO 276 protein ("H"; SEQ ID NO: 33) and the nucleotide sequences of the cDNA encoding murine protein M-Sema-F ("M"; SEQ ID NO: 66) is shown. These alignments were made using the

- 17 -

ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4).

Figure 4 comprises Figures 4A through 4M. The nucleotide sequence (SEQ ID NO: 38) of a cDNA encoding the human TANGO 292 protein described

5   herein is listed in Figures 4A to 4C. The ORF (residues 205 to 882; SEQ ID NO: 39) of the cDNA is indicated by nucleotide triplets, beneath which the amino acid sequence (SEQ ID NO: 40) of human TANGO 292 is listed. Figure 4D is a hydrophilicity plot of human TANGO 292 protein. The nucleotide sequence (SEQ ID NO: 81) of a cDNA encoding the gerbil TANGO 292 protein described herein is listed in Figures 4E to 4H.

10   The ORF (residues 89 to 763; SEQ ID NO: 82) of the cDNA is indicated by nucleotide triplets, below which the amino acid sequence (SEQ ID NO: 83) of gerbil TANGO 292 is listed. Figures 4I to 4K are an alignment of the nucleotide sequences of the ORF encoding human TANGO 292 protein ("H"; SEQ ID NO: 38) and the nucleotide sequence of the ORF encoding gerbil TANGO 292 protein ("G"; SEQ ID NO: 81),

15   made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4), wherein identical nucleotide residues are indicated by " | ". Figure 4L is an alignment of the human (H) and gerbil (G) TANGO 292 amino acid sequences, made using the same software and parameters, wherein identical amino acid residues are indicated by

20   " | " and similar amino acid residues are indicated by ".". Figure 4M is a hydrophilicity plot of gerbil TANGO 292 protein.

Figure 5 comprises Figures 5A through 5Mxviii. The nucleotide sequence (SEQ ID NO: 46) of a cDNA encoding the human TANGO 325 protein described herein is listed in Figures 5A through 5E. The ORF (residues 135 to 2000;

25   SEQ ID NO: 47) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 48) of human TANGO 325 is listed. Figure 5F is a hydrophilicity plot of TANGO 325 protein. An alignment of the amino acid sequences of TANGO 325 ("325"; SEQ ID NO: 48) and Slit-1 protein ("Slit"; SEQ ID NO: 67) protein is shown in Figures 5G to 5L. In Figures 5Mi to 5Mxviii, an alignment of the

nucleotide sequences of the cDNA encoding human TANGO 325 protein ("325"; SEQ ID NO: 33) and the nucleotide sequence of the cDNA encoding Slit-1 protein ("Slit"; SEQ ID NO: 68) is shown. This alignment was made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4).

Figure 6 comprises Figures 6A through 6J. The nucleotide sequence (SEQ ID NO: 54) of a cDNA encoding the human TANGO 331 protein described herein is listed in Figures 6A, 6B, and 6C. The ORF (residues 114 to 1172; SEQ ID NO: 55) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 56) of human TANGO 331 is listed. Figure 6D is a hydrophilicity plot of TANGO 331 protein. An alignment of the amino acid sequences of human TANGO 331 protein ("H"; SEQ ID NO: 56) and Chinese hamster protein HT ("C"; SEQ ID NO: 69; GenBank Accession No. U48852) is shown in Figure 6E. In Figures 6F through 6J, an alignment of the nucleotide sequences of the cDNA encoding human TANGO 331 protein ("H"; SEQ ID NO: 54) and the nucleotide sequence of the cDNA encoding Chinese hamster protein HT ("C"; SEQ ID NO: 70) is shown. These alignments were made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4).

Figure 7 comprises Figures 7A through 7U. The nucleotide sequence (SEQ ID NO: 59) of a cDNA encoding the human TANGO 332 protein described herein is listed in Figures 7A through 7E. The ORF (residues 173 to 2185; SEQ ID NO: 60) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 61) of human TANGO 332 protein is listed. Figure 7F is a hydrophilicity plot of TANGO 332 protein. An alignment of the amino acid sequences of TANGO 332 protein ("332"; SEQ ID NO: 61) and BEF protein ("BEF"; SEQ ID NO: 71) is shown in Figures 7G and 7H. An alignment of the amino acid sequences of human TANGO 332 protein ("H"; SEQ ID NO: 61) and murine brevidin protein ("M"; SEQ ID NO: 72) is shown in Figures 7I to 7K. In Figures 7L through 7U, an alignment

of the nucleotide sequences of the cDNA encoding human TANGO 332 protein ("H";
SEQ ID NO: 60) and the nucleotide sequence of the cDNA encoding murine brevidin
protein ("M"; SEQ ID NO: 73) is shown. These alignments were made using the
ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring
matrix; gap opening pentaly = 12, gap extension penalty = 4).

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of a
variety of human cDNA molecules which encode proteins which are herein designated
INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325,
TANGO 331, and TANGO 332. These proteins exhibit a variety of physiological
activities, and are included in a single application for the sake of convenience. It is
understood that the allowability or non-allowability of claims directed to one of these
proteins has no bearing on the allowability of claims directed to the others. The
characteristics of each of these proteins and the cDNAs encoding them are now
described separately.

INTERCEPT 217

A cDNA clone (designated jthqc035f08) encoding at least a portion of
human INTERCEPT 217 protein was isolated from a human prostate cDNA library.
The human INTERCEPT 217 protein is predicted by structural analysis to be a
transmembrane protein. In addition, cDNA clones (including those designated
jtmca047g07, jTmob373b05, and jambd078d12) encoding at least a portion of murine
INTERCEPT 217 protein were isolated from murine cDNA libraries.

The full length of the cDNA encoding human INTERCEPT 217 protein
(Figure 1; SEQ ID NO: 1) is 2895 nucleotide residues. The ORF of this cDNA,
nucleotide residues 215 to 1579 of SEQ ID NO: 1 (i.e., SEQ ID NO: 2), encodes a 455-
amino acid transmembrane protein (Figure 1; SEQ ID NO: 3). The murine ORF
(Figure 1; SEQ ID NO: 92) comprises at least 962 nucleotide residues. The protein

encoded by the murine ORF compises at least 320 amino acid residues (i.e., SEQ ID NO: 93), and is also a transmembrane protein.

The invention also includes purified human INTERCEPT 217 protein, both in the form of the immature 455 amino acid residue protein (SEQ ID NO: 3) and in the form of the mature, approximately 435 amino acid residue protein (SEQ ID NO: 5). Mature human INTERCEPT 217 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature INTERCEPT 217 protein and cleaving the signal sequence therefrom.

The invention thus includes purified murine INTERCEPT 217 protein, both in the immature form comprising the 320 amino acid residues of SEQ ID NO: 93 and in the mature form comprising the approximately 305 carboxyl terminal amino acid residues of SEQ ID NO: 93 (i.e., comprising SEQ ID NO: 95). Mature murine INTERCEPT 217 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature INTERCEPT 217 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature human and murine INTERCEPT 217 proteins, the invention includes fragments, derivatives, and variants of these INTERCEPT 217 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as INTERCEPT 217 polypeptides of the invention or INTERCEPT 217 proteins of the invention.

The invention also includes nucleic acid molecules which encode an INTERCEPT 217 polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 1, in SEQ ID NO: 92 (i.e., the murine ORF), or in some portion of either of these, such as the portion which encodes mature human INTERCEPT 217 protein, immature human INTERCEPT 217 protein, or a domain of human INTERCEPT 217 protein. These nucleic acids are collectively referred to as INTERCEPT 217 nucleic acids of the invention.

- 21 -

INTERCEPT 217 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features. Each of these molecules is included in the invention. As used herein, the term "family" is intended to mean two or more proteins or nucleic acid molecules

5   having a common or similar domain structure and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species. For example, a family can comprise two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin (e.g., the human and murine INTERCEPT 217 proteins described

10   herein).

A common domain present in INTERCEPT 217 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such

15   as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a

20   lipid bilayer. Thus, in one embodiment, a INTERCEPT 217 protein contains a signal sequence corresponding to about amino acid residues 1 to 20 of SEQ ID NO: 3 (SEQ ID NO: 4). The signal sequence is cleaved during processing of the mature protein.

INTERCEPT 217 proteins can include an extracellular domain. As used herein, an "extracellular domain" refers to a portion of a protein which is localized to

25   the non-cytoplasmic side of a lipid bilayer of a cell when a nucleic acid encoding the protein is expressed in the cell. The human INTERCEPT 217 protein extracellular domain is located from about amino acid residue 21 to about amino acid residue 383 of SEQ ID NO: 3 (SEQ ID NO: 6). The murine INTERCEPT 217 protein extracellular

- 22 -

domain is located from about amino acid residue 17 to about amino acid residue 213 of SEQ ID NO: 93 (SEQ ID NO: 96).

In addition, INTERCEPT 217 includes a transmembrane domain. As used herein, a "transmembrane domain" refers to an amino acid sequence which is at least about 20 to 25 amino acid residues in length and which contains at least about 65-70% hydrophobic amino acid residues such as alanine, leucine, phenylalanine, protein, tyrosine, tryptophan, or valine. In a preferred embodiment, a transmembrane domain contains at least about 15 to 30 amino acid residues, preferably about 20-25 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues. Thus, in one embodiment, an INTERCEPT 217 protein of the invention contains a transmembrane domain corresponding to about amino acid residues 384 to 403 of SEQ ID NO: 3 (SEQ ID NO: 7) or to about amino acid residues 214 to 233 of SEQ ID NO: 93 (SEQ ID NO: 97).

The present invention includes INTERCEPT 217 proteins having a cytoplasmic domain, particularly including proteins having a carboxyl-terminal cytoplasmic domain. As used herein, a "cytoplasmic domain" refers to a portion of a protein which is localized to the cytoplasmic side of a lipid bilayer of a cell when a nucleic acid encoding the protein is expressed in the cell. The human INTERCEPT 217 cytoplasmic domain is located from about amino acid residue 404 to amino acid residue 455 of SEQ ID NO: 3 (SEQ ID NO: 8). The murine INTERCEPT 217 cytoplasmic domain is located from aboud amino acid residue 234 to amino acid residue 320 of SEQ ID NO: 93 (SEQ ID NO: 98).

In one embodiment, the amino acid residues of human INTERCEPT 217 corresponding to SEQ ID NO: 8 are part of an extracellular domain, and the amino acid residues corresponding to SEQ ID NO: 6 are part of a cytoplasmic domain. In another embodiment, the amino acid residues of murine INTERCEPT 217 corresponding to SEQ ID NO: 98 are part of an extracellular domain, and the amino acid residues corresponding to SEQ ID NO: 96 are part of a cytoplasmic domain.

INTERCEPT 217 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Tables IA (for human INTERCEPT 217) and IB (for murine INTERCEPT 217), as predicted by computerized sequence analysis of INTERCEPT 217 proteins using amino acid sequence comparison software (comparing the amino acid sequence of INTERCEPT 217 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, or 10 or more of the post-translational modification sites listed in Tables IA and IB.

Table IA

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 3	Amino Acid Sequence
N-glycosylation site	107 to 110	NASG
	272 to 275	NCSS
	301 to 304	NTSV
	362 to 365	NQTH
	368 to 371	NVSV
Protein kinase C phosphorylation site	120 to 122	TLR
	192 to 194	SNR
	295 to 297	SLR
Casein kinase II phosphorylation site	199 to 202	SVPE
	440 to 443	TPPD
Tyrosine Kinase Phosphorylation Site	282 to 289	KRPEEHLY

N-myristoylation site	8 to 13	GTLLCM
	19 to 24	GTPDSE
	103 to 108	GVFVNA
	179 to 184	GLSATH
	323 to 328	GSRDGS
	348 to 353	GLFVCL
	390 to 395	GCAVGL
	449 to 454	GQASTS
Leucine zipper pattern	45 to 66	See Fig. 1
Leucine rich repeat amino terminal domain (LLRNT)	33 to 61	See Fig. 1
Leucine rich repeat (LRR) Domain	62 to 85	See Fig. 1
	86 to 109	See Fig. 1
	110 to 133	See Fig. 1
	134 to 157	See Fig. 1
	158 to 181	See Fig. 1
	184 to 207	See Fig. 1
Leucine rich repeat carboxyl terminal (LLRCT) domain	219 to 274	See Fig. 1

Table IB

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 93	Amino Acid Sequence
N-glycosylation site	102 to 105	NCSV
	131 to 134	NTSV
	192 to 195	NQTL
	198 to 201	NVSV
cAMP- and cGMP-dependent protein kinase site	280 to 283	RKAS
Protein kinase C phosphorylation site	125 to 127	SLR
	143 to 145	SPK
	279 to 281	SRK
Casein kinase II phosphorylation site	29 to 32	SIPE
	273 to 276	TPPD
N-myristoylation site	9 to 14	GLGLTR
	178 to 183	GVFVCL
	220 to 225	GCIVGL
	239 to 244	GCCHCC
Amidation Site	293 to 296	PGKK
Immunoglobulin Domain	14 to 37	See Fig. 1
Leucine rich repeat (LRR) Domain	49 to 104	See Fig. 1
Leucine rich repeat carboxyl terminal (LLRCT) domain	123 to 184	See Fig. 1

Among the domains that occur in INTERCEPT 217 proteins are LRR domains, LRRNT domains, LRRCT domains, and immunoglobulin domains. In one embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to one of these domains. In other embodiments, the protein has at least one of each of the LRR, LRRNT, and LRRCT domains described herein in Tables IA and IB. In other embodiments, the protein has at least one LRRNT domain, at least one LRRCT domain, and a plurality of (e.g., 2, 3, 4, or more) LRR domains.

One or more LRR domains are present in a variety of proteins involved in protein-protein interactions. Such proteins include, for example, proteins involved in signal transduction, cell-to-cell adhesion, cell-to-extracellular matrix adhesion, cell development, DNA repair, RNA processing, and cellular molecular recognition processes. Specialized LRR domains, designated LRR amino terminal (LRRNT) domains and LRR carboxyl terminal (LRRCT) domains often occur near the amino and carboxyl, respectively, ends of a series of LRR domains. Human INTERCEPT 217 protein has eight clustered LRR domains, including (from the amino terminus toward the carboxyl terminus of INTERCEPT 217) an LRRNT domain, six LRR domains, and an LRRCT domain.

The organization of LRR domains in human INTERCEPT 217 protein closely mirrors the organization of LRR domains in human platelet glycoprotein IB alpha chain precursor (GP-IBα), which also has eight clustered LRR domains from about amino acid residue 19 to about amino acid residue 281 thereof. The eight LRR domains of GP-IBα include an LRRNT domain at the end of the cluster nearest the amino terminus of GP-IBα and an LRRCT domain at the end of the cluster nearest the carboxyl terminus of GP-IBα. GP-IBα is a membrane-bound protein of human platelets that is involved in binding of von Willebrand's factor and in aggregation of platelets during thrombus formation. Thus, INTERCEPT 217 is involved in both normal and aberrant physiological activities involving blood clotting and thrombus

- 27 -

formation. Examples of disorders involving such activities include, for example, stroke, embolism (e.g., cerebral, renal, and pulmonary emboli), hemophilia, restenotic injury, prosthesis-associated thrombogenesis, atherosclerosis, and arteriosclerosis.

INTERCEPT 217 is involved in one or more physiological processes in which these other LRR domain-containing proteins are involved, namely binding of cells with extracellular proteins such as soluble extracellular proteins and cell surface proteins of other cells.

Human INTERCEPT 217 comprises a leucine zipper region at about amino acid residue 45 to about amino acid residue 66 (i.e., 45 LsctglgLqdvpaeLpaa tadL 66). Leucine zipper regions are known to be involved in dimerization of proteins. Leucine zipper regions interact with one another, leading to formation of homo- or hetero-dimers between proteins, depending on their identity. The presence in INTERCEPT 217 of a leucine zipper region is a further indication that this protein is involved in protein-protein interactions.

The amino acid sequence of human INTERCEPT 217 protein includes multiple potential proline-rich Src homology 3 (SH3) domain binding sites in the cytoplasmic portion of the protein. SH3 domains mediate specific assembly of protein complexes, presumably by interacting with proline-rich protein domains (Morton and Campbell (1994) *Curr. Biol.* 4:615-617). SH3 domains also mediate interactions between proteins involved in transmembrane signal transduction. Coupling of proteins mediated by SH3 domains has been implicated in a variety of physiological systems, including those involving regulation of cell growth and proliferation, endocytosis, and activation of respiratory burst.

SH3 domains have been described in the art (e.g., Mayer et al. (1988) *Nature* 332:272-275; Musacchio et al. (1992) *FEBS Lett.* 307:55-61; Pawson and Schlessinger (1993) *Curr. Biol.* 3:434-442; Mayer and Baltimore (1993) *Trends Cell Biol.* 3:8-13; Pawson (1993) *Nature* 373:573-580), and occur in a variety of cytoplasmic proteins, including several (e.g., protein tyrosine kinases) involved in transmembrane signal transduction. Among the proteins in which one or more SH3

domains occur are protein tyrosine kinases such as those of the Src, Abl, Bkt, Csk and ZAP70 families, mammalian phosphatidylinositol-specific phospholipases C-gamma-1 and -2, mammalian phosphatidylinositol 3-kinase regulatory p85 subunit, mammalian Ras GTPase-activating protein (GAP), proteins which mediate binding of guanine

5   nucleotide exchange factors and growth factor receptors (e.g., vertebrate GRB2, *Caenorhabditis elegans* sem-5, and *Drosophila* DRK proteins), mammalian Vav oncoprotein, guanidine nucleotide releasing factors of the CDC 25 family (e.g., yeast CDC25, yeast SCD25, and fission yeast ste6 proteins), MAGUK proteins (e.g., mammalian tight junction protein ZO-1, vertebrate erythrocyte membrane protein p55,

10   *C. elegans* protein lin-2, rat protein CASK, and mammalian synaptic proteins SAP90/PSD-95, CHAPSYN-110/PSD-93, SAP97/DLG1, and SAP102), proteins which interact with vertebrate receptor protein tyrosine kinases (e.g., mammalian cytoplasmic protein Nck and oncoprotein Crk), chicken Src substrate p80/85 protein (cortactin), human hemopoietic lineage cell specific protein Hs1, mammalian

15   dihydrouridine-sensitive L-type calcium channel beta subunit, human myasthenic syndrome antigen B (MSYB), mammalian neutrophil cytosolic activators of NADPH oxidase (e.g., p47 {NCF-1}, p67 {NCF-2}, and *C. elegans* protein B0303.7), myosin heavy chains (MYO3) from amoebae, from slime molds, and from yeast, vertebrate and *Drosophila* spectrin and fodrin alpha chain proteins, human amphiphysin, yeast actin-

20   binding proteins ABP1 and SLA3, yeast protein BEM1, fission yeast protein scd2 (ral3), yeast BEM1-binding proteins BOI2 (BEB1) and BOB1 (BOI1), yeast fusion protein FUS1, yeast protein RSV167, yeast protein SSU81, yeast hypothetical proteins YAR014c, YFR024c, YHL002w, YHR016c, YJL020C, and YHR114w, hypothetical fission yeast protein SpAC12C2.05c, and *C. elegans* hypothetical protein F42H10.3.

25   Of these proteins, multiple SH3 domains occur in vertebrate GRB2 protein, *C. elegans* sem-5 protein, *Drosophila* DRK protein, oncoprotein Crk, mammalian neutrophil cytosolic activators of NADPH oxidase p47 and p67, yeast protein BEM1, fission yeast protein scd2, yeast hypothetical protein YHR114w, mammalian cytoplasmic protein Nck, *C. elegans* neutrophil cytosolic activator of NADPH oxidase B0303.7, and yeast

actin-binding protein SLA1. Of these proteins, three or more SH3 domains occur in mammalian cytoplasmic protein Nck, *C. elegans* neutrophil cytosolic activator of NADPH oxidase B0303.7, and yeast actin-binding protein SLA1. The presence of SH3 domain binding sites in INTERCEPT 217 indicates that INTERCEPT 217 interacts

5   with one or more of these and other SH3 domain-containing proteins and is thus involved in physiological processes in which one or more of these or other SH3 domain-containing proteins are involved.

Human INTERCEPT 217 exhibits amino acid sequence similarity to porcine ribonuclease inhibitor, a protein which binds with high affinity to pancreatic

10   ribonucleases and inhibits their activity. INTERCEPT 217 thus is involved with similar physiological processes in humans. An alignment of the amino acid sequences of human INTERCEPT 217 and porcine ribonuclease inhibitor protein (SwissProt Accession number P10775) is shown in Figure 1G. In this alignment (made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring

15   matrix; gap opening pentaly = 12, gap extension penalty = 4), the proteins are 20.5% identical. An alignment of human (SEQ ID NO: 3) and murine INTERCEPT 217 amino acid sequences (SEQ ID NO: 93; made using BESTFIT software, BLOSUM62 scoring matrix, gap opening penalty = 12, frameshift gap penalty = 5, gap extension penalty = 4). In this alignment, the human and murine amino acid sequences are 71.3%

20   identical in the overlapping region. Alignment of human and murine INTERCEPT 217 ORFs indicated 79.9% nucleotide sequence identity in the overlapping region.

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human INTERCEPT 217 protein includes an approximately 20 (i.e., 18, 19, 20, 21, or 22) amino acid residue signal peptide

25   (amino acid residues 1 to 20 of SEQ ID NO: 3; SEQ ID NO: 4) preceding the mature INTERCEPT 217 protein (i.e., approximately amino acid residues 21 to 455 of SEQ ID NO: 3; SEQ ID NO: 5). In one embodiment, human INTERCEPT 217 protein includes an extracellular domain (amino acid residues 21 to 383 of SEQ ID NO: 3; SEQ ID NO: 6); a transmembrane domain (amino acid residues 384 to 403 of SEQ ID NO: 3; SEQ

ID NO: 7); and a cytoplasmic domain (amino acid residues 404 to 455 of SEQ ID NO:. 3; SEQ ID NO: 8). In an alternative embodiment, human INTERCEPT 217 protein includes a cytoplasmic domain (amino acid residues 21 to 383 of SEQ ID NO: 3; SEQ ID NO: 6); a transmembrane domain (amino acid residues 384 to 403 of SEQ ID NO:

5   3; SEQ ID NO: 7); and an extracellular domain (amino acid residues 404 to 455 of SEQ ID NO: 3; SEQ ID NO: 8).

The SIGNALP program predicted that murine INTERCEPT 217 protein includes an approximately 15 (i.e., 13, 14, 15, 16, or 17) amino acid residue signal peptide (amino acid residues 1 to 16 of SEQ ID NO: 93; SEQ ID NO: 94) preceding

10   the mature INTERCEPT 217 protein (i.e., approximately amino acid residues 16 to 320 of SEQ ID NO: 93; SEQ ID NO: 95). In one embodiment, murine INTERCEPT 217 protein includes an extracellular domain (amino acid residues 16 to 213 of SEQ ID NO: 93; SEQ ID NO: 96); a transmembrane domain (amino acid residues 214 to 233 of SEQ ID NO: 93; SEQ ID NO: 97); and a cytoplasmic domain (amino acid residues 234

15   to 320 of SEQ ID NO: 93; SEQ ID NO: 98). In an alternative embodiment, murine INTERCEPT 217 protein includes a cytoplasmic domain (amino acid residues 16 to 213 of SEQ ID NO: 93; SEQ ID NO: 96); a transmembrane domain (amino acid residues 214 to 233 of SEQ ID NO: 93; SEQ ID NO: 97); and an extracellular domain (amino acid residues 234 to 320 of SEQ ID NO: 93; SEQ ID NO: 98).

20   Figure 1F depicts a hydrophilicity plot of human INTERCEPT 217 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 20 of SEQ ID NO: 3 is the signal sequence of human INTERCEPT 217 (SEQ ID NO: 4). The hydrophobic region which

25   corresponds to amino acid residues 384 to 403 of SEQ ID NO: 3 is the transmembrane domain of human INTERCEPT 217 (SEQ ID NO: 7). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human INTERCEPT 217 protein

from about amino acid residue 355 to about amino acid residue 380 appears to be located at or near the surface of the protein, while the region from about amino acid residue 190 to about amino acid residue 210 appears not to be located at or near the surface. Figure 1L depicts a hydrophilicity plot of murine INTERCEPT 217 protein.

5   The predicted molecular weight of human INTERCEPT 217 protein without modification and prior to cleavage of the signal sequence is about 49.8 kilodaltons. The predicted molecular weight of the mature human INTERCEPT 217 protein without modification and after cleavage of the signal sequence is about 47.4 kilodaltons.

10   The predicted molecular weight of murine INTERCEPT 217 protein, without modification and prior to cleavage of the signal sequence is about 35.5 kilodaltons. The predicted molecular weight of the mature human INTERCEPT 217 protein without modification and after cleavage of the signal sequence is about 33.8 kilodaltons.

15   Northern analysis experiments indicated that mRNA corresponding to the cDNA encoding INTERCEPT 217 is expressed in two forms, one having an apparent approximate size of about 6 kilobases and another having an apparent approximate size of about 3 kilobases (i.e., corresponding to the size of the INTERCEPT 217 cDNA). These experiments indicated that INTERCEPT 217 is

20   expressed in the tissues listed in Table II, wherein "++" indicates strong expression, "+" indicates lower expression, and "+/-" indicates still lower expression.

Table II

Animal	Tissue	Relative Level of Expression
Human	pancreas	++
	skeletal muscle	+
	heart	+/-
	brain	+/-
	placenta	+/-
	lung	+/-
	liver	+/-
	kidney	+/-

An assay to detect possible secretion of INTERCEPT 217 protein was negative. This assay was performed as follows. About 8×10^5 293T cells were incubated at 37°C in wells containing growth medium (Dulbecco's modified Eagle's medium {DMEM} supplemented with 10% fetal bovine serum) under a 5% (v/v) CO_2, 95% air atmosphere to about 60-70% confluence. The cells were then transfected using a standard transfection mixture comprising 2 micrograms of DNA and 10 microliters of LIPOFECTAMINE™ (GIBCO/BRL Catalog no. 18342-012) per well. The transfection mixture was maintained for about 5 hours, and then replaced with fresh growth medium and maintained in an air atmosphere. Each well was gently rinsed twice with DMEM which did not contain methionine or cysteine (DMEM-MC; ICN Catalog no. 16-424-54). About 1 milliliter of DMEM-MC and about 50 microcuries of TRANS-^{35}S™ reagent (ICN Catalog no. 51006) were added to each well. The wells were maintained under the 5% CO_2 atmosphere described above and incubated at 37°C for a selected period. Following incubation, 150 microliters of conditioned medium was removed, centrifuged to remove floating cells and debris, and combined with 150 microliters of 2× SDS sample buffer. The sample was boiled at 100°C for 5 minutes, and about 40 microliters of sample was loaded onto a NOVEX™ 4-20% (w/v) SDS-

containing polyacrylamide gel. Following electrophoresis, the gel was stained for protein and dried according to the NOVEX™ procedure. The dried gel was exposed to radiation-sensitive film in order to detect the position of secreted proteins.

5   Biological function of INTERCEPT 217 proteins, nucleic acids encoding them, and modulators of these molecules

INTERCEPT 217 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that INTERCEPT 217 is expressed in pancreas,

10   skeletal muscle, heart, brain, placenta, lung, liver, and kidney tissue, INTERCEPT 217 protein is involved in one or more biological processes which occur in these tissues. In particular, INTERCEPT 217 is involved in modulating binding of cells of one or more of these tissues with proteins of other cells or with secreted proteins which occur in the extracellular environment of one or more of these tissues. INTERCEPT 217 is

15   especially implicated in disorders of skeletal muscle (e.g., protection of skeletal muscle cells during ischemia and in bruised tissue), and more especially those involving the pancreas (e.g., diabetes, pancreatitis, and the like).

Structural similarity of human INTERCEPT 217 protein with human GP-IBα indicates that INTERCEPT 217 is involved in binding extracellular proteins

20   and other ligands. INTERCEPT 217 protein is involved in binding of proteins which induce release of pancreatic digestive enzymes (e.g., amylases, lipases, proteases, and nucleases) from pancreatic cells, and in disorders associated with insufficient or inappropriate release of such enzymes. INTERCEPT 217 protein is also involved in binding of secreted pancreatic digestive enzymes in pancreatic tissue, thereby

25   protecting pancreatic tissue from autodigestion. Thus, INTERCEPT 217 protein is involved in disorders such as diabetes, pancreatitis, and pancreatic carcinoma which involve acute and chronic autodigestive damage to pancreatic tissues. Homology of INTERCEPT 217 protein with porcine ribonuclease inhibitor protein is a further indication of this involvement.

The presence of LRR domains in human INTERCEPT 217 protein and detection of its expression in a variety of tissues indicate that the tissue protective functions of INTERCEPT 217 are not limited to pancreatic tissues, but are involved in protection of other tissues as well (e.g., skeletal muscle, heart, brain, placenta, lung,

5   liver, prostate, and kidney tissues). INTERCEPT 217 is therefore involved in protection of these (and likely other tissues) from the effects of inflammation, autoimmunity, infection, and acute and chronic traumas.

Presence in INTERCEPT 217 protein of multiple SH3 domain binding sites indicates that INTERCEPT 217 protein interacts with one or more SH3 domain-

10   containing proteins. Thus, INTERCEPT 217 protein mediates binding of proteins (i.e., binding of proteins to INTERCEPT 217 and to one another to form protein complexes) in cells in which it is expressed. INTERCEPT 217 is also involved in transduction of signals between the exterior environment of cells (i.e., including from other cells) and the interior of cells in which it is expressed. INTERCEPT 217 mediates regulation of

15   cell growth and proliferation, endocytosis, activation of respiratory burst, and other physiological processes triggered by transmission of a signal via a protein with which INTERCEPT 217 interacts.

INTERCEPT 217-related molecules can be used to modulate one or more of the activities in which INTERCEPT 217 is involved and can also be used to

20   prevent, diagnose, or treat one or more of the disorders in which INTERCEPT 217 is involved.

INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof, can, for example, be used to treat pancreatic disorders, such as pancreatitis (e.g., acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (e.g., congenital

25   cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (e.g., pancreatic carcinoma and adenoma), diabetes mellitus (e.g., insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), and islet cell tumors (e.g., insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma). INTERCEPT 217 polypeptides, nucleic acids, and modulators

thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof, can be used to treat disorders of skeletal muscle, such as muscular dystrophy (e.g., Duchenne muscular dystrophy, Becker muscular dystrophy, Emery-Dreifuss muscular dystrophy, limb-girdle muscular dystrophy, facioscapulohumeral muscular dystrophy, myotonic dystrophy, oculopharyngeal muscular dystrophy, distal muscular dystrophy, and congenital muscular dystrophy), motor neuron diseases (e.g., amyotrophic lateral sclerosis, infantile progressive spinal muscular atrophy, intermediate spinal muscular atrophy, spinal bulbar muscular atrophy, and adult spinal muscular atrophy), myopathies (e.g., inflammatory myopathies {e.g., dermatomyositis and polymyositis}, myotonia congenita, paramyotonia congenita, central core disease, nemaline myopathy, myotubular myopathy, and periodic paralysis), and metabolic diseases of muscle (e.g., phosphorylase deficiency, acid maltase deficiency, phosphofructokinase deficiency, debrancher enzyme deficiency, mitochondrial myopathy, carnitine deficiency, carnitine palmityl transferase deficiency, phosphoglycerate kinase deficiency, phosphoglycerate mutase deficiency, lactate dehydrogenase deficiency, and myoadenylate deaminase deficiency). INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Because INTERCEPT 217 exhibits expression in heart tissue, INTERCEPT 217 nucleic acids, proteins, and modulators thereof can be used to treat heart disorders (e.g., ischemic heart disease, atherosclerosis, hypertension, angina pectoris, hypertrophic cardiomyopathy, and congenital heart disease). INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof, can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, e.g., infection, toxins, or

- 36 -

drugs), inflammations (e.g., bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (e.g., hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (e.g., neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain. INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof, can be used to treat placental disorders, such as toxemia of pregnancy (e.g., preeclampsia and eclampsia), placentitis, and spontaneous abortion. INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof, can be used to treat pulmonary (i.e., lung) disorders, such as atelectasis, cystic fibrosis, rheumatoid lung disease, pulmonary congestion, pulmonary edema, chronic obstructive airway disease (e.g., emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (e.g., sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), and tumors (e.g., bronchogenic carcinoma, bronchiolovlveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors). INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof, can be used to treat cardiovascular disorders, such as ischemic

heart disease (e.g., angina pectoris, myocardial infarction, and chronic ischemic heart disease), hypertensive heart disease, pulmonary heart disease, valvular heart disease (e.g., rheumatic fever and rheumatic heart disease, endocarditis, mitral valve prolapse, and aortic valve stenosis), congenital heart disease (e.g., valvular and vascular

5   obstructive lesions, atrial or ventricular septal defect, and patent ductus arteriosus), and myocardial disease (e.g., myocarditis, congestive cardiomyopathy, and hypertrophic cariomyopathy). INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

10   In yet another example, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof, can be used to treat hepatic (i.e., liver) disorders, such as jaundice, hepatic failure, hereditary hyperbiliruinemias (e.g., Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (e.g., hepatic vein thrombosis and portal vein obstruction and

15   thrombosis), hepatitis (e.g., chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis), cirrhosis (e.g., alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), and malignant tumors (e.g., primary carcinoma, hepatoblastoma, and angiosarcoma). INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more

20   of these disorders.

          In still another example, INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof, can be used to treat renal (i.e., kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis,

25   glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis),

tubulointerstitial diseases (e.g., pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy), acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia,

5   atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), and tumors (e.g., renal cell carcinoma and nephroblastoma). INTERCEPT 217 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

10   INTERCEPT 297

A cDNA clone (designated jthsa085g01) encoding at least a portion of human INTERCEPT 297 protein was isolated from a human fetal spleen cDNA library. The human INTERCEPT 297 protein is predicted by structural analysis to be a transmembrane protein.

15   The full length of the cDNA encoding human INTERCEPT 297 protein (Figure 2; SEQ ID NO: 9) is 1518 nucleotide residues. The ORF of this cDNA, nucleotide residues 40 to 1152 of SEQ ID NO: 9 (i.e., SEQ ID NO: 10), encodes a 371-amino acid transmembrane protein (Figure 2; SEQ ID NO: 11).

The invention thus includes purified human INTERCEPT 297 protein,

20   both in the form of a 371 amino acid residue protein (SEQ ID NO: 11) in which the 'signal sequence' (i.e., the portion of INTERCEPT 297 protein corresponding to amino acid residues 1 to 18) described in this section is not cleaved and in the form of a 353 amino acid residue protein (SEQ ID NO: 13) in which the 'signal sequence' is cleaved. Human INTERCEPT 297 protein can exist with or without the signal sequence

25   polypeptide at the amino terminus thereof. It is likely that the 'signal sequence' is not cleaved, but is instead a transmembrane domain of the protein.

In addition to full length human INTERCEPT 297 proteins, the invention includes fragments, derivatives, and variants of these INTERCEPT 297 proteins, as described herein. These proteins, fragments, derivatives, and variants are

collectively referred to herein as INTERCEPT 297 polypeptides of the invention or INTERCEPT 297 proteins of the invention.

The invention also includes nucleic acid molecules which encode an INTERCEPT 297 polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 9 or some portion thereof, such as the portion which encodes mature INTERCEPT 297 protein, immature INTERCEPT 297 protein, or a domain of INTERCEPT 297 protein. These nucleic acids are collectively referred to as INTERCEPT 297 nucleic acids of the invention.

INTERCEPT 297 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features.

A common domain present in INTERCEPT 297 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a INTERCEPT 297 protein contains a signal sequence corresponding to about amino acid residues 1 to 18 of SEQ ID NO: 11 (SEQ ID NO: 12). The signal sequence can be cleaved during processing of the mature protein, but it is likely that amino acid residues 1 to 18 of SEQ ID NO: 11 represent a (non-cleaved) transmembrane region of the protein.

INTERCEPT 297 proteins can include one or more extracellular domains. In one embodiment of the human INTERCEPT 297 protein, extracellular domains are located from about amino acid residues 19 to 47, from about amino acid

residues 110 to 118, from about amino acid residues 162 to 175, from about amino acid residues 234 to 260, and from about amino acid residues 313 to 319 of SEQ ID NO: 11 (SEQ ID NOs: 14-18, respectively). In an alternative embodiment, extracellullar domains are located from about amino acid residue 69 to 88, from about amino acid

5   residue 138 to 144, from about amino acid residue 193 to 215, from about amino acid residue 284 to 292, and from about amino acid residue 337 to 371 of SEQ ID NO: 11 (SEQ ID NOs: 28-32, respectively).

In addition, INTERCEPT 297 includes one or more transmembrane domains. In one embodiment, a INTERCEPT 297 protein of the invention contains

10   transmembrane domains corresponding to about amino acid residues 48 to 68, about amino acid residues 89 to 109, about amino acid residues 119 to 137, about amino acid residues 145 to 161, about amino acid residues 176 to 192, about amino acid residues 216 to 233, about amino acid residues 261 to 283, about amino acid residues 293 to 312, and about amino acid residues 320 to 336 of SEQ ID NO: 11 (SEQ ID NOs: 19-

15   27, respectively). As indicated above, it is likely that the 'signal sequence' of INTERCEPT 297 is an additional (and non-cleaved) transmembrane region.

The present invention includes INTERCEPT 297 proteins having one or more cytoplasmic domains. In one embodiment of the human INTERCEPT 297 protein, cytoplasmic domains are located from about amino acid residue 69 to 88, from

20   about amino acid residue 138 to 144, from about amino acid residue 193 to 215, from about amino acid residue 284 to 292, and from about amino acid residue 337 to 371 of SEQ ID NO: 11 (SEQ ID NOs: 28-32, respectively). In an alternative embodiment, cytoplasmic domains are located from about amino acid residues 19 to 47, from about amino acid residues 110 to 118, from about amino acid residues 162 to 175, from about

25   amino acid residues 234 to 260, and from about amino acid residues 313 to 319 of SEQ ID NO: 11 (SEQ ID NOs: 14-18, respectively).

INTERCEPT 297 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table III, as predicted by computerized sequence analysis of

- 41 -

INTERCEPT 297 proteins using amino acid sequence comparison software (comparing the amino acid sequence of INTERCEPT 297 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table III.

Table III

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 11	Amino Acid Sequence
N-glycosylation site	110 to 113	NMTS
	269 to 272	NISS
Protein kinase C phosphorylation site	24 to 26	SAK
	290 to 292	TTR
	297 to 299	SLR
Casein kinase II phosphorylation site	78 to 81	SSVD
	165 to 168	SKHD
	245 to 248	TLED
	354 to 357	SEQE
N-myristoylation site	18 to 23	GSINTL
	35 to 40	GCGGSK
	53 to 58	GMFLGE
	74 to 79	GQSDSS
	147 to 152	GILATI
	236 to 241	GSFSGN
	268 to 273	GNISSI
	280 to 285	GISVTK
Amidation site	136 to 139	LGRR
DUF6 domain	44 to 171	See Fig. 2

Among the domains that occur in INTERCEPT 297 protein is a DUF6 domain. In one embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to this DUF6 domain.

The DUF6 domain is a transmembrane domain that is highly conserved among eukaryote, prokaryote, and archae kingdoms. This high degree of domain sequence conservation indicates that proteins of the class which includes INTERCEPT 297 are involved in fundamental membrane physiology of living cells. INTERCEPT 297 protein is therefore involved in disorders which are associated with aberrant membrane function including, for example, disorders involving abnormal membrane fluidity, disorders involving aberrant transmembrane transport, disorders involving abnormal membrane organization, disorders involving abnormal membrane synthesis, disorders involving aberrant cell division, and the like.

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human INTERCEPT 297 protein includes an approximately 18 (i.e., 16, 17, 18, 19, or 20) amino acid residue signal peptide (amino acid residues 1 to 18 of SEQ ID NO: 11; SEQ ID NO: 12) preceding the mature INTERCEPT 297 protein (i.e., approximately amino acid residues 19 to 371 of SEQ ID NO: 11; SEQ ID NO: 13). In one embodiment, human INTERCEPT 297 protein includes about five extracellular domains (amino acid residues 19 to 47, 110 to 118, 162 to 175, 234 to 260, and 313 to 319 of SEQ ID NO: 11); about nine transmembrane domains (amino acid residues 48 to 68, 89 to 109, 119 to 137, 145 to 161, 176 to 192, 216 to 233, 261 to 283, 293 to 312, and 320 to 326 of SEQ ID NO: 11); and about five cytoplasmic domains (amino acid residues 69 to 88, 138 to 144, 193 to 215, 284 to 292, and 337 to 371 of SEQ ID NO: 11). In an alternative embodiment, human INTERCEPT 297 protein includes about five cytoplasmic domains (amino acid residues 19 to 47, 110 to 118, 162 to 175, 234 to 260, and 313 to 319 of SEQ ID NO:

11); about nine transmembrane domains (amino acid residues 48 to 68, 89 to 109, 119 to 137, 145 to 161, 176 to 192, 216 to 233, 261 to 283, 293 to 312, and 320 to 326 of SEQ ID NO: 11); and about five extracellular domains (amino acid residues 69 to 88, 138 to 144, 193 to 215, 284 to 292, and 337 to 371 of SEQ ID NO: 11).

5   Figure 2D depicts a hydrophilicity plot of human INTERCEPT 297 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. Hydrophobic region corresponding to the signal sequence and the transmembrane domains are observed in this figure. As described elsewhere herein, relatively hydrophilic regions 10 are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human INTERCEPT 297 protein from about amino acid residue 165 to about amino acid residue 175 appears to be located at or near the surface of the protein.

The predicted molecular weight of human INTERCEPT 297 protein 15 without modification and prior to cleavage of the signal sequence is about 40.2 kilodaltons. The predicted molecular weight of the mature human INTERCEPT 297 protein without modification and after cleavage of the signal sequence is about 38.2 kilodaltons.

20   <u>Biological function of INTERCEPT 297 proteins, nucleic acids encoding them, and modulators of these molecules</u>

INTERCEPT 297 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that INTERCEPT 297 is expressed in human fetal 25 spleen, INTERCEPT 297 protein is involved in one or more biological processes which occur in fetal and spleen tissues. In particular, INTERCEPT 297 is involved in modulating growth, proliferation, survival, differentiation, and activity of cells including, but not limited to, spleen and fetal cells of the animal in which it is normally expressed. Thus, INTERCEPT 297 has a role in disorders which affect these cells and

their growth, proliferation, survival, differentiation, and activity (e.g., hematologic and immune disorders). Expression of INTERCEPT 297 in an animal is also involved in modulating growth, proliferation, survival, differentiation, and activity of cells and viruses which are foreign to the host (i.e., bacterial, fungal, and viral infections).

5   INTERCEPT 297 bears amino acid sequence similarity to Caenorhabditis elegans protein C2G12.12, and therefore exhibits one or more activities analogous to that protein.

INTERCEPT 297 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, differentiation, or function of cells of the spleen

10   (e.g., cells of the splenic connective tissue, splenic smooth muscle cells, and endothelial cells of the splenic blood vessels). INTERCEPT 297 nucleic acids, proteins, and modulators thereof can also be used to modulate the proliferation, differentiation, and function of cells that are processed within the spleen (e.g., regenerated or phagocytized within the spleen, erythrocytes, B and T lymphocytes, and

15   macrophages). Thus, INTERCEPT 297 nucleic acids, proteins, and modulators thereof can be used to treat disorders of the spleen (including disorders of the fetal spleen). Examples of splenic disorders include, splenic lymphoma, splenomegaly, and phagocytotic disorders (e.g., those in which macrophage engulfment of bacteria and viruses in the bloodstream is inhibited). INTERCEPT 297 polypeptides, nucleic acids,

20   and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Structural analysis of INTERCEPT 297 and the presence of a DUF6 domain therein indicate that INTERCEPT 297 is involved in disorders which affect membrane structure and function. INTERCEPT 297 can be used to affect development

25   and persistence of disorders involving inappropriate membrane structure and function, such as atherogenesis, arteriosclerosis, and various transmembrane transport disorders. Other exemplary disorders for which INTERCEPT 297 is useful include disorders involving generation and persistence of an immune response to bacterial, fungal, and viral infections. INTERCEPT 297 polypeptides, nucleic acids, and modulators thereof

can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

The structure of INTERCEPT 297 is analogous to the structures of integral membrane proteins responsible for transmembrane transport of molecules such as sugars, ions, and the like. INTERCEPT 297 is thus involved in one or more transmembrane transport-related disorders such as cystic fibrosis, nerve conduction disorders (e.g., pain and loss or failure of sensation), muscle contraction disorders (e.g., cardiac insufficiency), metal ion uptake disorders (e.g., hemochromatosis), and the like. INTERCEPT 297 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

TANGO 276

A cDNA clone (designated jthsa006e01) encoding at least a portion of human TANGO 276 protein was isolated from a human fetal spleen cDNA library. The human TANGO 276 protein is predicted by structural analysis to be a secreted protein.

The full length of the cDNA encoding human TANGO 276 protein (Figure 3; SEQ ID NO: 33) is 2811 nucleotide residues. The ORF of this cDNA, nucleotide residues 58 to 786 of SEQ ID NO: 33 (i.e., SEQ ID NO: 34), encodes a 243-amino acid secreted protein (Figure 3; SEQ ID NO: 35).

The invention thus includes purified human TANGO 276 protein, both in the form of the immature 243 amino acid residue protein (SEQ ID NO: 35) and in the form of the mature, approximately 223 amino acid residue protein (SEQ ID NO: 37). Mature human TANGO 276 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature TANGO 276 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature human TANGO 276 proteins, the invention includes fragments, derivatives, and variants of these TANGO 276 proteins, as described herein. These proteins, fragments, derivatives, and variants

are collectively referred to herein as TANGO 276 polypeptides of the invention or TANGO 276 proteins of the invention.

The invention also includes nucleic acid molecules which encode a TANGO 276 polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 33 or some portion thereof, such as the portion which encodes mature TANGO 276 protein, immature TANGO 276 protein, or a domain of TANGO 276 protein. These nucleic acids are collectively referred to as TANGO 276 nucleic acids of the invention.

TANGO 276 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features, as indicated by the conservation of amino acid sequence between human TANGO 276 protein and the murine protein designated M-Sema-F (see Inagaki et al. (1995) *FEBS Lett.* 370:269-272), as shown in Figures 3F to 3H.

A common domain present in TANGO 276 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 276 protein contains a signal sequence corresponding to about amino acid residues 1 to 20 of SEQ ID NO: 35 (SEQ ID NO: 36). The signal sequence is cleaved during processing of the mature protein.

TANGO 276 proteins can exist in a secreted form, such as a mature protein having the amino acid sequence of amino acid residues 21 to 243 of SEQ ID NO: 35 (SEQ ID NO: 37).

- 47 -

TANGO 276 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table IV, as predicted by computerized sequence analysis of TANGO 276 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 276 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, or all 8 of the post-translational modification sites listed in Table IV.

Table IV

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 35	Amino Acid Sequence
N-glycosylation site	106 to 109 121 to 124	NQTE NASH
cAMP- or cGMP-dependent protein kinase phosphorylation site	43 to 46	RRFS
Protein kinase C phosphorylation site	194 to 196	SLK
Casein kinase II phosphorylation site	34 to 37 57 to 60	SSGE TLTE
N-myristoylation site	16 to 21 68 to 73	GLGIGA GAREAL
Sema domain	53 to 141	See Fig. 3

A Sema domain occurs in human TANGO 276 protein. In one embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to this Sema domain.

- 48 -

Sema domains occur in semaphorin proteins. Semaphorins are a large family of secreted and transmembrane proteins, some of which function as repellent signals during neural axon guidance. The Sema domain and a variety of semaphorin proteins in which it occurs are described, for example, in Winberg et al. (1998 *Cell*

5   95:903-916). Sema domains also occur in human hepatocyte growth factor receptor (SwissProt Accession no. P08581) and the similar neuronal and epithelial transmembrane receptor protein (SwissProt Accession no. P51805). The presence of a Sema domain in human TANGO 276 protein indicates that TANGO 276 is involved in one or more physiological processes in which the semaphorins are involved, has

10   biological activity in common with one or more of the semaphorins, or both.

Human TANGO 276 protein exhibits considerable sequence similarity to murine M-Sema F protein (GenBank Accession no. S79463), as indicated herein in Figures 3F to 3H. Figures 3F to 3H depict an alignment of the amino acid sequences of human TANGO 276 protein (SEQ ID NO: 35) and murine M-Sema F protein (SEQ ID

15   NO: 65). In this alignment (pam120.mat scoring matrix, gap opening pentaly = 12, gap extension penalty = 4), the amino acid sequences of the proteins are 76.1% identical. Figures 3I through 3R depict an alignment of the nucleotide sequences of cDNA encoding human TANGO 276 protein (SEQ ID NOs: 33) and murine cDNA encoding M-Sema F protein (SEQ ID NO: 66). In this alignment (pam120.mat scoring matrix,

20   gap opening pentaly = 12, gap extension penalty = 4), the nucleic acid sequences of the cDNAs are 79.7% identical. Thus, TANGO 276 is related to murine M-Sema F and shares functional similarities to that protein.

It is known that semaphorins are bi-functional, capable of functioning either as attractive axonal guidance proteins or as repellent axonal guidance proteins

25   (Wong et al. (1997) *Development* 124:3597-3607). Furthermore, semaphorins bind with neuronal cell surface proteins designated plexins, which are expressed on both neuronal cells and cells of the immune system (Comeau et al. (1998) *Immunity* 8:473-482; Jin and Strittmatter (1997) *J. Neurosci.* 17:6256-6263).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 276 protein includes an approximately 20 (i.e., 18, 19, 20, 21, or 22) amino acid signal peptide (amino acid residues 1 to 20 of SEQ ID NO: 35; SEQ ID NO: 36) preceding the mature TANGO 276 protein (i.e., approximately amino acid residues 21 to 243 of SEQ ID NO: 34; SEQ ID NO: 37). Human TANGO 276 protein is a secreted protein.

Figure 3E depicts a hydrophilicity plot of human TANGO 276 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to about amino acid residues 1 to 20 of SEQ ID NO: 35 is the signal sequence of human TANGO 276. As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 276 protein from about amino acid residue 90 to about amino acid residue 105 appears to be located at or near the surface of the protein, while the region from about amino acid residue 170 to about amino acid residue 180 appears not to be located at or near the surface.

The predicted molecular weight of human TANGO 276 protein without modification and prior to cleavage of the signal sequence is about 27.1 kilodaltons. The predicted molecular weight of the mature human TANGO 276 protein without modification and after cleavage of the signal sequence is about 24.8 kilodaltons.

Northern analysis experiments indicated that mRNA corresponding to the cDNA encoding TANGO 276 is expressed in the tissues listed in Table V, wherein "++" indicates a greater level of expression and "+" indicates a lower level of expression.

Table V

Animal	Tissue	Relative Level of Expression
Human	heart	++
	placenta	++
	brain	+
	lung	+
	liver	+
	skin	+
	kidney	+
	pancreas	+

Biological function of TANGO 276 proteins, nucleic acids encoding them, and modulators of these molecules

TANGO 276 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that TANGO 276 is expressed in human heart and placenta tissues, to a lesser extent in brain, lung, liver, skin, kidney, and pancreas tissues, and in fetal spleen tissue, TANGO 276 protein is involved in one or more biological processes which occur in these tissues. In particular, TANGO 276 is involved in modulating growth, proliferation, survival, differentiation, and activity of cells including, but not limited to, heart, placenta, spleen, brain, lung, liver, skin, kidney, and pancreas cells of the animal in which it is normally expressed. Thus, TANGO 276 has a role in disorders which affect these cells and their growth, proliferation, survival, differentiation, and activity.

Because TANGO 276 exhibits expression in the heart, TANGO 276 nucleic acids, proteins, and modulators thereof can be used to treat heart disorders. Examples of heart disorders with which TANGO 276 can be involved include ischemic heart disease, atherosclerosis, hypertension, angina pectoris, hypertrophic

cardiomyopathy, and congenital heart disease. TANGO 276 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, TANGO 276 polypeptides, nucleic acids, and modulators thereof can be used to treat placental disorders, such as toxemia of pregnancy (e.g., preeclampsia and eclampsia), placentitis, and spontaneous abortion. TANGO 276 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, TANGO 276 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, e.g., infection, toxins, or drugs), inflammations (e.g., bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (e.g., hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (e.g., neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain. TANGO 276 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

TANGO 276 polypeptides, nucleic acids, and modulators thereof can be associated with pulmonary (i.e., lung) disorders, such as atelectasis, cystic fibrosis, rheumatoid lung disease, pulmonary congestion, pulmonary edema, chronic obstructive airway disease (e.g., emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (e.g., sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), and tumors (e.g., bronchogenic carcinoma,

bronchiolovlveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors). TANGO 276 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, TANGO 276 polypeptides, nucleic acids, and modulators thereof, can be used to treat hepatic (i.e., liver) disorders, such as jaundice, hepatic failure, hereditary hyperbiliruinemias (e.g., Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (e.g., hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis (e.g., chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (e.g., alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), and malignant tumors (e.g., primary carcinoma, hepatoblastoma, and angiosarcoma). TANGO 276 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Exemplary skin disorders with which TANGO 276 can be associated include, by way of example, psoriasis, infections, wounds (and healing of wounds), inflammation, dermatitis, acne, benign and malignant dermatological tumors, and the like. TANGO 276 proteins, nucleic acids encoding them, and agents that modulate activity or expression of either of these can be used to prognosticate, diagnose, treat, and inhibit one or more of these disorders.

In another example, TANGO 276 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal (i.e., kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (e.g., pyelonephritis, drug and toxin induced tubulointerstitial

nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), and tumors (e.g., renal cell carcinoma and nephroblastoma). TANGO 276 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Pancreatic disorders in which TANGO 276 can be involved include pancreatitis (e.g., acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (e.g., congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (e.g., pancreatic carcinoma and adenoma), diabetes mellitus (e.g., insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), and islet cell tumors (e.g., insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma). TANGO 276 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

The presence of the Sema domain in TANGO 276 indicates that this protein is involved in development of neuronal and epithelial tissues and also functions as a repellant protein which guides axonal development. TANGO 276 modulates nerve growth and regeneration and also modulates growth and regeneration of other epithelial tissues. TANGO 276 is thus involved in a variety of neuronal disorder including, but not limited to, one or more of seizure, epilepsy, (regeneration of) neuronal damage, pain (including, for example, migraine, headache, and other chronic pain), infections of the central nervous system, multiple sclerosis, sleep disorders, psychological disorders, nerve root disorders, and the like. Presence of a Sema domain in TANGO 276 further indicates that TANGO 276 has one or more physiological roles in common with other proteins (e.g., secreted and transmembrane semaphorins, collapsins, neuropilins, plexins, and the like) in which the Sema domain occurs. Thus, TANGO 276 is implicated in development, maintenance, and regeneration of neuronal connections and

networks, in modulating differentiation of cells of the immune system, in modulating

cytokine production by cells of the immune system, in modulating reactivity of cells of

the immune system toward cytokines, in modulating initiation and persistence of an

inflammatory response, and in modulating proliferation of epithelial cells. Sema

5   domain-containing proteins have also been implicated in development and progression

of small cell lung cancer, in normal brain development, and immune system regulation.

This indicates that TANGO 276 is also involved in one or more of these processes and

in disorders relating to these processes (e.g., small cell lung cancer, brain development

disorders, and immune and auto-immune disorders). TANGO 276 polypeptides,

10   nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit,

prevent, or alleviate one or more of these disorders.

        The observation that TANGO 276 shares identity with the murine

semaphorin protein designated M-Sema F suggests that TANGO 276 has activity

identical or analogous to the activity of this protein. These observations indicate that

15   TANGO 276 modulates growth, proliferation, survival, differentiation, and activity of

neuronal cells. Thus, TANGO 276 protein is useful, for example, for modulating and

guiding neural axon development and for modulating establishment and maintenance

of neuronal networks.

20   <u>TANGO 292</u>

        A cDNA clone (designated jthkf040b11) encoding at least a portion of

human TANGO 292 protein was isolated from a human normal embryonic keratinocyte

cDNA library. A corresponding gerbil cDNA clone (designated jtiba040e12) was also

isolated, and encoded at least a portion of gerbil TANGO 292 protein. The human and

25   TANGO 292 proteins are predicted by structural analysis to be transmembrane

proteins.

        The full length of the cDNA encoding human TANGO 292 protein

(Figure 4; SEQ ID NO: 38) is 2498 nucleotide residues. The ORF of this cDNA,

nucleotide residues 205 to 882 of SEQ ID NO: 38 (i.e., SEQ ID NO: 39), encodes a

226-amino acid residue transmembrane protein (Figure 4; SEQ ID NO: 40). The full

length of the cDNA encoding gerbil TANGO 292 protein (Figure 4; SEQ ID NO: 81) is

2002 nucleotide residues. The ORF of this cDNA, nucleotide residues 89 to 763 of

SEQ ID NO: 81 (i.e., SEQ ID NO: 82), encodes a 225-amino acid transmembrane

5   protein (Figure 4; SEQ ID NO: 83).

The invention thus includes purified human TANGO 292 protein, both

in the form of the immature 226 amino acid residue protein (SEQ ID NO: 40) and in

the form of the mature, approximately 209 amino acid residue protein (SEQ ID NO:

42). The invention also includes purified gerbil TANGO 292 protein, both in the form

10   of the immature 225-amino acid residue (SEQ ID NO: 83) protein and in the form of

the mature, approximately 208-amino acid residue protein (SEQ ID NO: 85). Mature

human or gerbil TANGO 292 protein can be synthesized without the signal sequence

polypeptide at the amino terminus thereof, or it can be synthesized by generating

immature TANGO 292 protein and cleaving the signal sequence therefrom.

15   In addition to full length mature and immature human and gerbil

TANGO 292 proteins, the invention includes fragments, derivatives, and variants of

these TANGO 292 proteins, as described herein. These proteins, fragments,

derivatives, and variants are collectively referred to herein as TANGO 292

polypeptides of the invention or TANGO 292 proteins of the invention.

20   The invention also includes nucleic acid molecules which encode a

TANGO 292 polypeptide of the invention. Such nucleic acids include, for example, a

DNA molecule having the nucleotide sequence listed in SEQ ID NO: 38 or 81 or some

portion thereof, such as the portion which encodes mature human or gerbil TANGO

292 protein, immature human or gerbil TANGO 292 protein, or a domain of human or

25   gerbil TANGO 292 protein. These nucleic acids are collectively referred to as TANGO

292 nucleic acids of the invention.

TANGO 292 proteins and nucleic acid molecules encoding them

comprise a family of molecules having certain conserved structural and functional

features. This family includes, for example, human and gerbil TANGO 292 proteins and nucleic acid molecules described herein.

A common domain present in TANGO 292 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 292 protein contains a signal sequence corresponding to about amino acid residues 1 to 17 of SEQ ID NO: 40 (SEQ ID NO: 41) or to about amino acid residues 1 to 17 of SEQ ID NO: 83 (SEQ ID NO: 84). The signal sequence is cleaved during processing of the mature protein.

TANGO 292 proteins can include an extracellular domain. The human TANGO 292 protein extracellular domain is located from about amino acid residue 18 to about amino acid residue 113 of SEQ ID NO: 40 (SEQ ID NO: 43). The gerbil TANGO 292 protein extracellular domain includes at least about amino acid residues 18 to 112 of SEQ ID NO: 83 (SEQ ID NO: 86).

In addition, TANGO 292 include a transmembrane domain. In one embodiment, a human TANGO 292 protein contains a transmembrane domain corresponding to about amino acid residues 114 to 138 of SEQ ID NO: 40 (SEQ ID NO: 44). Gerbil TANGO 292 protein includes a transmembrane domain corresponding to about amino acid residues 113 to 137 of SEQ ID NO: 83 (SEQ ID NO: 87).

The present invention includes TANGO 292 proteins having a cytoplasmic domain, particularly including proteins having a carboxyl-terminal cytoplasmic domain. The human TANGO 292 cytoplasmic domain is located from about amino acid residue 139 to amino acid residue 226 of SEQ ID NO: 40 (SEQ ID

NO: 45). The gerbil TANGO 292 cytoplasmic domain is located from about amino acid residue 138 to amino acid residue 225 of SEQ ID NO: 83 (SEQ ID NO: 88).

TANGO 292 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table VIa as predicted by computerized sequence analysis of human TANGO 292 protein, or in Table VIb as predicted by computerized sequence analysis of gerbil TANGO 292 protein, using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 292 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, or all of the post-translational modification sites listed in Table VIa or in Table VIb.

Table VIa

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 40	Amino Acid Sequence
cAMP- or cGMP-dependent protein kinase phosphorylation site	197 to 200	RKHS
Protein kinase C phosphorylation site	37 to 39	TSK
	97 to 99	SAK
	102 to 104	TTK
	196 to 198	TRK
Casein kinase II phosphorylation site	37 to 40	TSKE
	103 to 106	TKSD
	180 to 183	SVED
N-myristoylation site	116 to 121	GLLTGL
Vitamin K-dependent carboxylation domain	56 to 98	See Fig. 4

Table VIb

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 83	Amino Acid Sequence
cAMP- or cGMP-dependent protein kinase phosphorylation site	196 to 199	RKHS
Protein kinase C phosphorylation site	23 to 25	SLK
	37 to 39	SKK
	96 to 98	SVK
	101 to 103	TTR
	155 to 157	TRR
	195 to 197	TRK
Casein kinase II phosphorylation site	74 to 77	SYEE
	102 to 105	TRSD
	155 to 157	THEE
	195 to 197	SSSE
N-myristoylation site	33 to 38	GVFASK
	115 to 120	GLLTGL
Vitamin K-dependent carboxylation domain	55 to 92	See Fig. 4

Among the domains that occur in TANGO 292 protein is a vitamin K-dependent carboxylation domain. In one embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to this vitamin K-dependent carboxylation domain.

The vitamin K-dependent carboxylation domain has the following consensus sequence, wherein standard single-letter amino acid codes are used and 'X' refers to any amino acid residue.

-X_{12}-E-X_3-E-X-C-X_6-(D or E or N)-X-(L or I or V or M or F or Y)-X_9-(F or Y or W)-

5   Glutamic acid residues within this consensus region are potential vitamin K-dependent carboxylation sites. Human TANGO 292 has 9 glutamic acid residues in the vitamin K-dependent carboxylation domain located from about amino acid residue 56 to 98 of SEQ ID NO: 40, namely at amino acid residues 58, 66, 68, 71, 72, 77, 78, 81, and 86 of SEQ ID NO: 40, and gerbil TANGO 292 has 10 glutamic acid residues in the vitamin

10   K-dependent carboxylation domain located from about amino acid residue 55 to 92 of SEQ ID NO: 83, namely at amino acid residues 57, 65, 67, 70, 71, 76, 77, 80, 86, and 87 of SEQ ID NO: 83. In one embodiment, the protein of the invention is carboxylated at one or more of these glutamic acid residues. In some proteins in which a vitamin K-dependent carboxylation domain occurs, many of the glutamic acid residues which

15   occur from the amino terminus of the protein through the conserved aromatic residue at the carboxyl terminal end of the domain are carboxylated. Human TANGO 292 has 13 glutamic acid residues in the region from the amino terminus of (both the immature and mature forms of) the protein and the tryptophan residue at amino acid residue 93 of SEQ ID NO: 40, and also has another glutamic acid residue at position 95 of SEQ ID

20   NO: 40 which can also be carboxylated. In addition, human TANGO 292 protein has four sets of paired (i.e., adjacent) glutamic acid residues, at residues 33-34, 40-41, 71-72, and 77-78 and a pair of glutamic acid residues (66 and 68) which are separated by a single residue. Similarly, gerbil TANGO 292 has 12 glutamic acid residues in the region from the amino terminus of (both the immature and mature forms of) the protein

25   and the tryptophan residue at amino acid residue 92 of SEQ ID NO: 83, and also has another glutamic acid residue at position 94 of SEQ ID NO: 83 which can also be carboxylated. In addition, gerbil TANGO 292 protein has three sets of glutamic acid residues, at residues 70-71, 76-77, and 86-87, and a pair of glutamic acid residues (65 and 67) which are separated by a single residue. The protein of the invention includes

proteins which are carboxylated at one or more of the individual or paired glutamic acid residues.

TANGO 292, like other vitamin K-dependent carboxylation domain-containing proteins, is involved in binding, uptake, and response to metal cations such as calcium, to proteins, and to small molecules. Other proteins in which a vitamin K-dependent carboxylation domain occurs include, for example, osteocalcin (bone-Gla protein), matrix Gla protein, various plasma proteins such as prothrombin, coagulation factors VII, IX, and X, proline rich Gla domain-containing proteins PRGP1 and PRGP2, and proteins C, S, and Z. Thus, TANGO 292 is involved in physiological processes in which one or more of these other vitamin K-dependent carboxylation domain-containing proteins is involved.

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 292 protein includes an approximately 17 (i.e., 15, 16, 17, 18, or 19) amino acid residue signal peptide (amino acid residues 1 to 17 of SEQ ID NO: 40; SEQ ID NO: 41) preceding the mature TANGO 292 protein (i.e., approximately amino acid residues 18 to 226 of SEQ ID NO: 40; SEQ ID NO: 42). In one embodiment, human TANGO 292 protein includes an extracellular domain (amino acid residues 18 to 113 of SEQ ID NO: 40; SEQ ID NO: 43); a transmembrane domain (amino acid residues 114 to 138 of SEQ ID NO: 40; SEQ ID NO: 44); and a cytoplasmic domain (amino acid residues 139 to 225 of SEQ ID NO: 40; SEQ ID NO: 45). In an alternative embodiment, human TANGO 292 protein includes a cytoplasmic domain (amino acid residues 18 to 113 of SEQ ID NO: 40; SEQ ID NO: 43); a transmembrane domain (amino acid residues 114 to 138 of SEQ ID NO: 40; SEQ ID NO: 44); and an extracellular domain (amino acid residues 139 to 225 of SEQ ID NO: 40; SEQ ID NO: 45).

The SignalP program predicted that gerbil TANGO 292 protein includes an approximately 17 (i.e., 15, 16, 17, 18, or 19) amino acid residue amino acid signal peptide (amino acid residues 1 to 17 of SEQ ID NO: 83; SEQ ID NO: 84) preceding the mature TANGO 292 protein (i.e., approximately amino acid residues 18 to 225 of

SEQ ID NO: 83; SEQ ID NO: 85). In one embodiment, gerbil TANGO 292 protein includes an extracellular domain (amino acid residues 18 to 112 of SEQ ID NO: 83; SEQ ID NO: 86); a transmembrane domain (amino acid residues 113 to 137 of SEQ ID NO: 83; SEQ ID NO: 87); and a cytoplasmic domain (amino acid residues 138 to 225 of SEQ ID NO: 83; SEQ ID NO: 88). In an alternative embodiment, gerbil TANGO 292 protein includes a cytoplasmic domain (amino acid residues 18 to 112 of SEQ ID NO: 83; SEQ ID NO: 86); a transmembrane domain (amino acid residues 113 to 137 of SEQ ID NO: 83; SEQ ID NO: 87); and an extracellular domain (amino acid residues 138 to 225 of SEQ ID NO: 83; SEQ ID NO: 88).

Figure 4E depicts a hydrophilicity plot of human TANGO 292 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 17 of SEQ ID NO: 40 is the signal sequence of human TANGO 292. The hydrophobic region which corresponds to amino acid residues 114 to 138 of SEQ ID NO: 40 is the transmembrane domain of human TANGO 292. As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 292 protein from about amino acid residue 90 to about amino acid residue 110 appears to be located at or near the surface of the protein, while the region from about amino acid residue 190 to about amino acid residue 195 appears not to be located at or near the surface.

Figure 4M depicts a hydrophilicity plot of gerbil TANGO 292 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 17 of SEQ ID NO: 83 is the signal sequence of gerbil TANGO 292. The hydrophobic region which corresponds to amino acid residues 113 to 137 of SEQ ID NO: 40 is the transmembrane domain of gerbil TANGO 292. As described elsewhere herein, relatively hydrophilic regions are

generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of gerbil TANGO 292 protein from about amino acid residue 90 to about amino acid residue 110 appears to be located at or near the surface of the protein.

5   An alignment of the human (H) and gerbil (G) ORF sequences encoding TANGO 292 protein is shown in Figures 4I-4K. This alignment was made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4), and indicates about 64.1% identity between these two cDNA sequences. An alignment of the amino acid

10   sequences of gerbil (G) and human (H) TANGO 292 proteins is shown in Figure 4L. In this alignment (made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4), the proteins are about 77.7% identical and about 80% similar.

          The predicted molecular weight of human TANGO 292 protein without

15   modification and prior to cleavage of the signal sequence is about 25.4 kilodaltons. The predicted molecular weight of the mature human TANGO 292 protein without modification and after cleavage of the signal sequence is about 23.6 kilodaltons. The predicted molecular weight of gerbil TANGO 292 protein without modification and prior to cleavage of the signal sequence is about 25.4 kilodaltons. The predicted

20   molecular weight of the mature human TANGO 292 protein without modification and after cleavage of the signal sequence is about 23.5 kilodaltons.

          Northern analysis experiments indicated that human mRNA corresponding to the cDNA encoding TANGO 292 is expressed in the tissues listed in Table VIc, wherein "++" indicates strong expression, "+" indicates lower expression,

25   "+/-" indicates still lower expression, and "-" indicates that expression could not be detected in the corresponding tissue.

Table VIc

Animal	Tissue	Relative Level of Expression
Human	placenta	++
	liver	++
	kidney	++
	lung	+
	pancreas	+
	heart	+/-
	brain	-
	skeletal muscle	-

<u>Biological function of TANGO 292 proteins, nucleic acids encoding them, and modulators of these molecules</u>

5   TANGO 292 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that TANGO 292 is expressed in human embryonic keratinocytes, and in placenta, liver, kidney, lung, pancreas, and heart tissues, TANGO 292 protein is involved in one or more biological processes which

10   occur in these tissues. In particular, TANGO 292 is involved in modulating growth, proliferation, survival, differentiation, and activity of cells including, but not limited to, keratinocytes and cells with which keratinocytes interact in the animal in which TANGO 292 is normally expressed. TANGO 292 is also involved in modulating growth, proliferation, survival, differentiation, and activity of placenta, liver, kidney,

15   lung, pancreas, and heart cells. Thus, TANGO 292 has a role in disorders which affect these cells and their growth, proliferation, survival, differentiation, and activity. TANGO 292 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, TANGO 292 polypeptides, nucleic acids, and

modulators thereof can be used to treat placental disorders, such as toxemia of pregnancy (e.g., preeclampsia and eclampsia), placentitis, and spontaneous abortion. TANGO 292 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

5   In another example, TANGO 292 polypeptides, nucleic acids, and modulators thereof, can be used to treat hepatic (i.e., liver) disorders, such as jaundice, hepatic failure, hereditary hyperbiliruinemias (e.g., Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (e.g., hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis

10   (e.g., chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (e.g., alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), and malignant tumors (e.g., primary carcinoma, hepatoblastoma, and angiosarcoma). TANGO 292 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

15   In another example, TANGO 292 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal (i.e., kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus

20   erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (e.g., pyelonephritis, drug and toxin induced tubulointerstitial

25   nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), and tumors (e.g., renal cell carcinoma and nephroblastoma). TANGO 292 polypeptides, nucleic acids, and

modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

TANGO 292 polypeptides, nucleic acids, and modulators thereof can be associated with pulmonary (i.e., lung) disorders, such as atelectasis, cystic fibrosis, rheumatoid lung disease, pulmonary congestion, pulmonary edema, chronic obstructive airway disease (e.g., emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (e.g., sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), and tumors (e.g., bronchogenic carcinoma, bronchiolovlveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors). TANGO 292 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Pancreatic disorders in which TANGO 292 can be involved include pancreatitis (e.g., acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (e.g., congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (e.g., pancreatic carcinoma and adenoma), diabetes mellitus (e.g., insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), and islet cell tumors (e.g., insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma). TANGO 292 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Because TANGO 292 exhibits expression in the heart, TANGO 292 nucleic acids, proteins, and modulators thereof can be used to treat heart disorders. Examples of heart disorders with which TANGO 292 can be involved include ischemic heart disease, atherosclerosis, hypertension, angina pectoris, hypertrophic cardiomyopathy, and congenital heart disease. TANGO 292 polypeptides, nucleic

acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Presence in TANGO 292 of a vitamin K-dependent carboxylation (Gla) domain indicates that TANGO 292 is involved in physiological functions identical or analogous to the functions performed by other proteins having such domains. For example, like other Gla domain-containing proteins, TANGO 292 modulates binding and uptake of calcium and other metal ions by cells which express it and the response of those cells to the presence and uptake of such ions. Human matrix Gla protein, for example, is involved in Keutel syndrome, an autosomal recessive disorder characterized by abnormal cartilage calcification, peripheral pulmonary stenosis, and midfacial hypoplasia (Munroe et al. (1999) *Nat. Genet.* 21:142-144). Other proteins containing a Gla domain include, for example, two human proline-rich Gla proteins designated PRGP1 and PRGP2, human G domain-containing protein Gas6, and several human blood coagulation factors (Kulman et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:9058-9062; Mark et al., (1996) *J. Biol. Chem.* 271:9785-9786; Cancela et al. (1990) *J. Biol. Chem.* 265:15040-15048). These proteins are involved in binding of mineral ions such as calcium, phosphate, and hydroxyapatite, binding of proteins, binding of vitamins and small molecules, and mediation of blood coagulation. Thus, TANGO 292 is involved in numerous physiological processes which are influenced by levels of calcium and other metal ions in body fluids or by the presence of proteins, vitamins, or small molecules. Such processes include, for example, bone uptake, maintenance, and deposition, formation, maintenance, and repair of cartilage, formation and maintenance of extracellular matrices, movement of cells through extracellular matrices, coagulation and dissolution of blood components (e.g., blood cells and proteins), and deposition of materials (e.g., lipids, cells, calcium, and the like) in arterial walls. TANGO 292 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

TANGO 292 is involved in disorders which affect the tissues in which it is normally expressed and upon which it normally acts. Thus, TANGO 292 is involved

in disorders which involve aberrant binding or aberrant failure to bind of keratinocytes or similar cells with a tissue affected by the disorder. Such disorders include, by way of example and not limitation, osteoporosis, (repair of) traumatic bone injuries, rickets, osteomalacia, Paget's disease, and other bone disorders, osteoarthritis, rheumatoid arthritis, ankylosing spondylitis, Keutel syndrome, and other disorders of the joints and cartilage, iron deficiency anemia, hemophilia, inappropriate blood coagulation, stroke, arteriosclerosis, atherosclerosis, aneurysm, and other disorders related to blood and blood vessels, metastasis and other disorders related to inappropriate movement of cells through extracellular matrices, and the like. TANGO 292 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

TANGO 325

A cDNA clone (designated jthdc071a12) encoding at least a portion of human TANGO 325 protein was isolated from a human aortic endothelial cell cDNA library. The human TANGO 325 protein is predicted by structural analysis to be a transmembrane protein.

The full length of the cDNA encoding human TANGO 325 protein (Figure 5; SEQ ID NO: 46) is 2169 nucleotide residues. The ORF of this cDNA, nucleotide residues 135 to 2000 of SEQ ID NO: 46 (i.e., SEQ ID NO: 47), encodes a 622-amino acid transmembrane protein (Figure 5; SEQ ID NO: 48).

The invention thus includes purified human TANGO 325 protein, both in the form of the immature 622 amino acid residue protein (SEQ ID NO: 48) and in the form of the mature, approximately 591 amino acid residue protein (SEQ ID NO: 50). Mature human TANGO 325 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature TANGO 325 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature human TANGO 325 proteins, the invention includes fragments, derivatives, and variants of these TANGO

325 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as TANGO 325 polypeptides of the invention or TANGO 325 proteins of the invention.

The invention also includes nucleic acid molecules which encode a TANGO 325 polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 46 or some portion thereof, such as the portion which encodes mature TANGO 325 protein, immature TANGO 325 protein, or a domain of TANGO 325 protein. These nucleic acids are collectively referred to as TANGO 325 nucleic acids of the invention.

TANGO 325 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features.

A common domain present in TANGO 325 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 325 protein contains a signal sequence corresponding to about amino acid residues 1 to 31 of SEQ ID NO: 48 (SEQ ID NO: 49). The signal sequence is cleaved during processing of the mature protein.

TANGO 325 proteins can include an extracellular domain. The human TANGO 325 protein extracellular domain is located from about amino acid residue 32 to about amino acid residue 529 of SEQ ID NO: 48 (SEQ ID NO: 51).

In addition, TANGO 325 include a transmembrane domain. In one embodiment, a TANGO 325 protein of the invention contains a transmembrane domain

corresponding to about amino acid residues 530 to 547 of SEQ ID NO: 48 (SEQ ID NO: 52).

The present invention includes TANGO 325 proteins having a cytoplasmic domain, particularly including proteins having a carboxyl-terminal cytoplasmic domain. The human TANGO 325 cytoplasmic domain is located from about amino acid residue 548 to amino acid residue 622 of SEQ ID NO: 48 (SEQ ID NO: 53).

TANGO 325 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table VII, as predicted by computerized sequence analysis of TANGO 325 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 325 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table VII.

Table VII

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 48	Amino Acid Sequence
N-glycosylation site	71 to 74	NISY
	76 to 79	NESE
	215 to 218	NLTK
	266 to 269	NVTR
	317 to 320	NDTF
	331 to 334	NLSF
	336 to 339	NLTA
	400 to 403	NITN
	410 to 413	NVSR
	451 to 454	NITF
	579 to 582	NVTA
cAMP- or cGMP-dependent protein kinase phosphorylation site	231 to 234	RRLS
Protein kinase C phosphorylation site	40 to 42	TGR
	229 to 231	SLR
	326 to 328	SLK
	390 to 392	SMR
	510 to 512	SGK
	575 to 577	SAR
Casein kinase II phosphorylation site	284 to 287	SHND
	442 to 445	SPLE
	447 to 450	TETE
	453 to 456	TFWE

Table VII (Continued)

N-myristoylation site	3 to 8	GLQFSL
	69 to 74	GNNISY
	126 to 131	GIFKGL
	174 to 179	GTFVGM
ATP/GTP-binding site motif A (P-loop)	506 to 513	AASMSGKT
Leucine rich repeat amino terminal domain (LLRNT)	32 to 60	See Fig. 5
Leucine rich repeat (LRR) domain	61 to 84	See Fig. 5
	85 to 108	See Fig. 5
	109 to 132	See Fig. 5
	133 to 156	See Fig. 5
	157 to 180	See Fig. 5
	181 to 204	See Fig. 5
	205 to 228	See Fig. 5
	229 to 252	See Fig. 5
	253 to 276	See Fig. 5
	277 to 300	See Fig. 5
	301 to 324	See Fig. 5
	326 to 349	See Fig. 5
Leucine rich repeat carboxyl terminal domain (LRRCT)	359 to 405	See Fig. 5

Among the domains that occur in TANGO 325 protein are leucine rich
repeat (LRR) domains, including amino terminal and carboxyl terminal LRR domains,
and a P-loop domain. In one embodiment, the protein of the invention has at least one

domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to one of these domains. In another embodiment, the protein has at least on amino terminal LRR domain, at least one carboxyl terminal LRR domain, and a

5   plurality of LRR domains interposed therebetween. In yet another embodiment, the protein has at least one P-loop domain, and a plurality (e.g., 2, 3, 4, or more) of the LRR domains described herein in Table VII.

One or more LRR domains is present in a variety of proteins involved in protein-protein interactions. Such proteins include, for example, proteins involved in

10   signal transduction, cell-to-cell adhesion, cell-to-extracellular matrix adhesion, cell development, DNA repair, RNA processing, and cellular molecular recognition processes. Specialized LRR domains, designated LRR amino terminal (LRRNT) domains and LRR carboxyl terminal (LRRCT) domains often occur near the amino and carboxyl, respectively, ends of a series of LRR domains. TANGO 325 protein has

15   fourteen clustered LRR domains, including (from the amino terminus toward the carboxyl terminus of TANGO 325) an LRRNT domain, twelve LRR domains, and an LRRCT domain. TANGO 325 is thus involved in one or more physiological processes in which these other LRR domain-containing proteins are involved, namely binding of cells with extracellular proteins such as soluble extracellular proteins and cell surface

20   proteins of other cells.

The fact that TANGO 325 has an ATP/GTP-binding domain (i.e., a P-loop domain) within the extracellular domain of the protein indicates that this protein is involved in transmembrane signaling events. Considered in combination with the protein-binding LRR domains present in the extracellular domain of the, the presence

25   of the ATP/GTP-binding domain indicates that TANGO 325 protein is capable of sensing extracellular proteins, including ATP-binding proteins and GTP-binding proteins, and extracellular nucleotides (e.g., ATP, ADP, and AMP). Thus, TANGO 325 protein is involved in translating information (e.g., environmental conditions or signaling molecules provided to the environment by other cells) from the extracellular

- 73 -

environment of the cell in which it is expressed to one or more intracellular biochemical systems.

TANGO 325 exhibits amino acid sequence and nucleic acid sequence homology with human Slit-1 protein. An alignment of the amino acid sequences of TANGO 325 and human Slit-1 protein is shown in Figures 5G to 5L. In this alignment (made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4), the proteins are 35.4% identical (i.e., 35.4% of the residues of TANGO 325 correspond to identical residues in Slit-1). An alignment of the nucleotide sequences of the ORFs encoding TANGO 325 and human Slit-1 protein is shown in Figures 5Mi through 5Mxviii. The two ORFs are 65.7% identical, as assessed using the same software and parameters.

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 325 protein includes an approximately 31 (i.e., 29, 30, 31, 32, or 33) amino acid residue signal peptide (amino acid residues 1 to 31 of SEQ ID NO: 48; SEQ ID NO: 49) preceding the mature TANGO 325 protein (i.e., approximately amino acid residues 42 to 622 of SEQ ID NO: 48; SEQ ID NO: 50). In one embodiment, human TANGO 325 protein includes an extracellular domain (amino acid residues 32 to 529 of SEQ ID NO: 48; SEQ ID NO: 51); a transmembrane domain (amino acid residues 530 to 547 of SEQ ID NO: 48; SEQ ID NO: 52); and a cytoplasmic domain (amino acid residues 548 to 622 of SEQ ID NO: 48; SEQ ID NO: 53). In an alternative embodiment, human TANGO 325 protein includes a cytoplasmic domain (amino acid residues 32 to 529 of SEQ ID NO: 48; SEQ ID NO: 51); a transmembrane domain (amino acid residues 530 to 547 of SEQ ID NO: 48; SEQ ID NO: 52); and an extracellular domain (amino acid residues 548 to 622 of SEQ ID NO: 48; SEQ ID NO: 53).

Figure 5F depicts a hydrophilicity plot of human TANGO 325 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region

which corresponds to amino acid residues 1 to 31 of SEQ ID NO: 48 is the signal sequence of human TANGO 325 (SEQ ID NO: 49). The hydrophobic region which corresponds to amino acid residues 530 to 547 of SEQ ID NO: 48 is the transmembrane domain of human TANGO 325 (SEQ ID NO: 52). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 325 protein from about amino acid residue 550 to about amino acid residue 565 appears to be located at or near the surface of the protein, while the region from about amino acid residue 168 to about amino acid residue 185 appears not to be located at or near the surface.

The predicted molecular weight of human TANGO 325 protein without modification and prior to cleavage of the signal sequence is about 70.3 kilodaltons. The predicted molecular weight of the mature human TANGO 325 protein without modification and after cleavage of the signal sequence is about 66.8 kilodaltons.

Northern analysis experiments indicated that mRNA corresponding to the cDNA encoding TANGO 325 is expressed in the tissues listed in Table VIIA, wherein "+" indicates expression and "-" indicates that expression could not be detected in the corresponding tissue.

Table VIIA

Animal	Tissue	Relative Level of Expression
Human	placenta	+
	liver	+
	kidney	+
	pancreas	+
	heart	+
	brain	-
	skeletal muscle	-
	lung	-

<u>Biological function of TANGO 325 proteins, nucleic acids encoding them, and</u>
<u>modulators of these molecules</u>

TANGO 325 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that TANGO 325 is expressed in human aortic endothelial tissue and in placenta, liver, kidney, pancreas, and heart tissues, TANGO 325 protein is involved in one or more biological processes which occur in these tissues. In particular, TANGO 325 is involved in modulating growth, proliferation, survival, differentiation, and activity of endothelial cells including, but not limited to, vascular and cardiac (including valvular) endothelial cells of the animal in which it is normally expressed. TANGO 325 also modulates growth, proliferation, survival, differentiation, and activity of placenta, liver, kidney, and pancreas cells. Thus, TANGO 325 has a role in disorders which affect these cells and their growth, proliferation, survival, differentiation, and activity. TANGO 325 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In one example, TANGO 325 polypeptides, nucleic acids, and modulators thereof can be used to treat placental disorders, such as toxemia of pregnancy (e.g., preeclampsia and eclampsia), placentitis, and spontaneous abortion. TANGO 325 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, TANGO 325 polypeptides, nucleic acids, and modulators thereof, can be used to treat hepatic (i.e., liver) disorders, such as jaundice, hepatic failure, hereditary hyperbiliruinemias (e.g., Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (e.g., hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis (e.g., chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (e.g., alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis),

- 76 -

and malignant tumors (e.g., primary carcinoma, hepatoblastoma, and angiosarcoma). TANGO 325 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, TANGO 325 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal (i.e., kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (e.g., pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), and tumors (e.g., renal cell carcinoma and nephroblastoma). TANGO 325 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Pancreatic disorders in which TANGO 325 can be involved include pancreatitis (e.g., acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (e.g., congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (e.g., pancreatic carcinoma and adenoma), diabetes mellitus (e.g., insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), and islet cell tumors (e.g., insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma). TANGO 325 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Because TANGO 325 exhibits expression in the heart, TANGO 325 nucleic acids, proteins, and modulators thereof can be used to treat heart disorders. Examples of heart disorders with which TANGO 325 can be involved include ischemic heart disease, atherosclerosis, hypertension, angina pectoris, hypertrophic

5   cardiomyopathy, and congenital heart disease. TANGO 325 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

It is known that serum nucleotide levels (e.g., ATP) affect cardiac contractility and vasomotor tone. Presence in TANGO 325 of an ATP/GTP binding

10   domain in the extracellular portion thereof implicates this transmembrane protein in sensing of serum nucleotide levels and transmission of the sensed level by mechanisms not yet fully understood to myocytes underlying the epithelium. Thus, TANGO 325 is involved in disorders such as cardiovascular insufficiency, hypertension, hypotension, shock, and the like.

15   Leukocytes are known to bind with vascular endothelial surfaces in a reversible manner prior to penetrating the vascular endothelium in route to an underlying tissue. Although a few proteins have previously been implicated in the leukocyte-endothelium binding process, the identities of all of the proteins involved remain unknown. The presence of numerous LRR domains on the exterior portion of

20   TANGO 325 protein implicates this protein in reversible binding of leukocytes to vascular endothelium. Thus, TANGO 325 is involved in physiological processes and disorders which involve leukocyte-endothelium binding. Such processes and disorders include, by way of example, cellular aspects of immune responses, autoimmune responses and disorders, and migration of leukocytes to lymph nodes.

25   The aortic endothelium, as well as other vascular endothelia, are known to be involved in detection of signals (e.g., metabolites, proteins, and the like) in the blood stream. Mammalian Slit-1 protein is known to be involved in the human endocrine system (Itoh et al. (1998) *Brain Res. Mol. Brain Res.* 62:175-186). Amino acid and nucleic acid sequence similarity of TANGO 325 with human Slit-1 protein, as

described herein, indicates that TANGO 325 is involved in sensing physiological signals by the endocrine system. Thus, TANGO 325 is involved in one or more human endocrine disorders such as pituitary disorders (e.g., diabetes insipidus), thyroid disorders (e.g., hyperthyroidism, hypothyroidism, diabetes, goiter, and growth and developmental disorders), adrenal disorders (e.g., Addison's disease, Cushing's syndrome, hyperaldosteronism, and pheochromocytoma), and the like.

Human Slit-1 protein is also known to be involved in guidance of neuronal growth. The sequence similarity of TANGO 325 with Slit-1, as described herein, implicates TANGO 325 in growth, development, maintenance, and regeneration of neurons. TANGO 325 can thus be used to prevent, diagnose, and treat a variety of neurological disorders.

TANGO 331

A cDNA clone (designated jthvb042g08) encoding at least a portion of human TANGO 331 protein was isolated from a human mammary epithelium cDNA library. A corresponding cDNA clone (designated jchrc045a03) was isolated from a human heart library. The human TANGO 331 protein is predicted by structural analysis to be a secreted protein.

The full length of the cDNA encoding human TANGO 331 protein (Figure 6; SEQ ID NO: 54) is 1432 nucleotide residues. The ORF of this cDNA, nucleotide residues 114 to 1172 of SEQ ID NO: 54 (i.e., SEQ ID NO: 55), encodes a 353-amino acid secreted protein (Figure 6; SEQ ID NO: 56).

The invention thus includes purified human TANGO 331 protein, both in the form of the immature 353 amino acid residue protein (SEQ ID NO: 56) and in the form of the mature, approximately 329 amino acid residue protein (SEQ ID NO: 58). Mature human TANGO 331 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature TANGO 331 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature human TANGO 331 proteins, the invention includes fragments, derivatives, and variants of these TANGO 331 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as TANGO 331 polypeptides of the invention or

5   TANGO 331 proteins of the invention.

The invention also includes nucleic acid molecules which encode a TANGO 331 polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 54 or some portion thereof, such as the portion which encodes mature TANGO 331 protein,

10   immature TANGO 331 protein, or a domain of TANGO 331 protein. These nucleic acids are collectively referred to as TANGO 331 nucleic acids of the invention.

TANGO 331 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features, as indicated by the conservation of amino acid sequence between human

15   TANGO 331 protein and the Chinese hamster (*Cricetulus griseus*) protein designated HT and having GenBank Accession number U48852, as shown in Figure 6E, and the conservation of nucleotide sequence between the ORFs encoding human TANGO 331 protein and Chinese hamster protein HT, as shown in Figures 6F through 6J.

A common domain present in TANGO 331 proteins is a signal

20   sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino

25   acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 331 protein contains a signal

sequence corresponding to about amino acid residues 1 to 24 of SEQ ID NO: 56 (SEQ ID NO: 57). The signal sequence is cleaved during processing of the mature protein.

TANGO 331 proteins can include an extracellular domain. The human TANGO 331 protein is a secreted protein, and thus includes an 'extracellular domain'
5 consisting of the entire mature protein (i.e., approximately residues 25 to 353 of SEQ ID NO: 56; SEQ ID NO: 58).

TANGO 331 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table VIII, as predicted by computerized sequence analysis of
10 TANGO 331 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 331 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table VIII.

Table VIII

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 56	Amino Acid Sequence
N-glycosylation site	190 to 193	NETH
	251 to 254	NGSY
cAMP- or cGMP-dependent protein kinase phosphorylation site	26 to 29	KKPT
Protein kinase C phosphorylation site	48 to 50	TAK
	123 to 125	TLK
	144 to 146	SQR
	165 to 167	SCR
	187 to 189	SLR
	202 to 204	SCK
	210 to 212	TNR
Casein kinase II phosphorylation site	58 to 61	TAWE
	66 to 69	SKYE
	86 to 89	SDFE
	197 to 200	TACD
	210 to 213	TNRD
	255 to 258	TCEE
	295 to 298	SLAE
	339 to 342	TEGE
	349 to 352	SRED

Table VIII (Continued)

Tyrosine kinase phosphorylation site	303 to 309	RKNENCY
N-myristoylation site	44 to 49	GMVDTA
	54 to 59	GGGNTA
	81 to 86	GLCESS
	150 to 155	GNGHCS
	158 to 163	GSRQGD
	164 to 169	GSCRCH
	252 to 257	GSYTCE
	313 to 318	GSYVCV
Aspartic acid and asparagine hydroxylation site	308 to 319	See Fig. 6
EGF-like domain cystein pattern signature	166 to 177	See Fig. 6
EGF domain	140 to 177	See Fig. 6
	234 to 263	See Fig. 6
	301 to 330	See Fig. 6
Laminin-like EGF domain	153 to 199	See Fig. 6
TNFR/NGFR cysteine-rich region domain	180 to 214	See Fig. 6
Vertebrate metallothionein-like domain	229 to 298	See Fig. 6
Leucine Zipper domain	94 to 115	See Fig. 6

Among the domains that occur in TANGO 331 protein are EGF domains, including a laminin-like EGF domain, a TNFR/NGFR cysteine-rich domain, a metallothionein-like domain, and a leucine zipper domain.

EGF-like domains are about 30 to 40 amino acid residues in length and comprise several conserved cysteine residues in one of several patterns. EGF-like domains occur in a large number of proteins including, for example, human epidermal growth factor (EGF), murine adipocyte differentiation inhibitor, human agrin, human growth factor amphiregulin, human growth factor betacellulin, sea urchin blastula tissue patterning proteins BP10 and Span, cattle tick glycoprotein BM86, human bone morphogenic protein 1, sea urchin suBMP, *Drosophila* tolloid protein, *Caenorhabditis elegans* developmental proteins lin-12 and glp-1, *C. elegans* tissue patterning protein APX-1, human calcium-dependent serine proteinase, human cartilage matrix protein, human cartilage oligomeric matrix protein, human cell surface antigen 114/A10, rat cell surface glycoprotein complex transmembrane subunit ASGP-2, human coagulation associated proteins C, Z, and S, human coagulation factors VII, IX, X, and XII, human complement components C1r, C1s, C6, C7, C8α, C8β, and C9, human complement-activating components of Ra-reactive factor, *Drosophila* epithelial development protein Crumbs, sea urchin exogastrula-inducing peptides A, C, D, and X, *Drosophila* cadherin-related tumor suppressor protein Fat, human fetal antigen 1 (a neuroendocrine differentiation protein derived from the delta-like protein), human fibrillins 1 and 2, sea urchin fibropellins IA, IB, IC, II, and III, human extracellular matrix proteins fibulin-1 and -2, *Drosophila* cell determination/axon guidance protein Argos, various poxvirus growth factor-related proteins, *Drosophila* developmental protein Gurken, human heparin-binding EGF-like growth factor, human transforming growth factor-α, human growth factors Lin-3 and Spitz, human hepatocyte growth factor activator, human LDL and VLDL receptors, human LDL receptor-related protein, human leukocyte antigen CD97, human cell surface glycoprotein EMR1, human cell surface glycoprotein F4/80, Japanese horseshoe crab limulus clotting factor C, mammalian membrane-bound endopeptidase Meprin A α subunit, murine milk fat globule-EGF factor 8, human glial

growth factors neuregulin GGF-I and GGF-II, mammalian neurexins, human neurogenic proteins Notch, Xotch, Tan-1, and Delta, *C. elegans* differentiation protein Lag-2, *Drosophila* differentiation proteins Serrate and Slit, chordate basement membrane protein Nidogen, *Plasmodium* ookinete 24, 25, and 28 kilodalton surface

5   proteins, human pancreatic secretory granule membrane glycoprotein GP2, human non-specific cell lysis protein Perforin, human proteoglycans aggrecan, versican, perlecan, brevican, and chondroitin sulfate, human endoplasmic reticulum prostaglandin G/H synthases 1 and 2, human extracellular protein S1-5, human autocrine growth factor Schwannoma-derived growth factor, human E-, P-, and L-selectins, *Arabidopsis*

10   *thaliana* chlorophyll complex assembly protein serine/threonine-protein kinase homolog, guinea pig sperm-egg fusion proteins PH-30α and β, murine stromal cell derived protein-1, human teratocarcinoma-derived growth factor, mammalian extracellular protein tenascin, chicken extracellular protein TEN-A, human tenascin-X, *Drosophila* tenascin-like proteins TEN-A and TEN-M, human protein C activator

15   thrombomodulin, human adhesive glycoproteins thrombospondins 1, 2, 3, and 4, human thyroid peroxidases 1 and 2, human transforming growth factor β-1 binding protein, human tyrosine-protein kinase receptors Tek and Tie, human urokinase-type plasminogen activator, human tissue plasminogen activator, human uromodulin, human vitamin K-dependent anticoagulant proteins C and S (and the related human single-

20   chain plasma glycoprotein Z), the sea urchin 63 kilodalton sperm flagellar membrane protein, chicken Nel protein, and the hypothetical *C. Elegans* protein T20G5.3. Although these proteins have a variety of activities and sites of expression, a common characteristic of most of them is that they are involved in protein-to-protein binding in the extracellular space - either to a secreted protein, a component of the extracellular

25   matrix, or to an extracellular portion of an integral membrane protein. Based on this shared characteristic, the presence of multiple EGF-like domains in TANGO 331 indicates that TANGO 331 is involved in binding to proteins extracellularly.

    Post-translational hydroxylation of aspartic acid or asparagine to form erythro-β-hydroxyaspartic acid or erythro-β-hydroxyasparagine occurs in various

- 85 -

proteins having one or more EGF-like domains (e.g., blood coagulation protein factors VII, IX, and X, blood coagulation proteins C, S, and Z, the LDL receptor, thrombomodulin, and the like). TANGO 331 has a signature sequence which is characteristic of hydroxylation of the asparagine residue at amino acid residue 310. The invention thus includes TANGO 331 proteins having a hydroxylated asparagine residue at position 310 of SEQ ID NO: 56.

TNFR/NGFR (tumor necrosis factor receptor/nerve growth factor receptor) cysteine-rich region domains are about 30 to 40 amino acid residues in length, and generally exhibit a conserved pattern of six or more cysteine residues. These domains occur in several soluble and transmembrane proteins which are known to be receptors for growth factors or for cytokines. Examples of TNFR/NGFR cysteine-rich region domain-containing proteins are human tumor necrosis factor (TNF) cysteine-rich region domains type I and type II receptors, Shope fibroma virus soluble TNF receptor, human lymphotoxin α/β, human low-affinity nerve growth factor receptor, human CD40L (cytokine) receptor CD40, human CD27L (cytokine) receptor CD27, human CD30L (cytokine) receptor CD30, human T-cell cytokine receptor 4-1BB, human apoptotic FASL protein receptor FAS, human T-cell OX40L (cytokine) receptor OX40, human apoptosis-related receptor Wsl-1, and *Vaccinia* protein A53. Presence of a TNFR/NGFR cysteine-rich region domain in TANGO 331 is an indication that TANGO 331 is involved in one or more physiological processes involving extracellular binding with a cytokine or growth factor. Such processes include, for example, growth, homeostasis, regeneration, and proliferation of cells and tissues, immune (including autoimmune) responses, host defenses against infection, and the like.

Metallothioneins are cysteine-rich proteins which are capable of binding heavy metals such as calcium, zinc, copper, cadmium, cobalt, nickel, and the like. Proteins which have a domain which resembles a metal-binding domain of a metallothionein are also capable of binding such metals. TANGO 331 comprises a metallothionein-like domain, and is capable of binding one or more heavy metals. This is an indication that TANGO 331 is involved in one or more physiological processes

- 86 -

which involve metal binding. Such processes include, by way of example and not limitation, nutritional supply of metals to cells on a controlled basis, removal of toxic metal species from body tissues, storage of metals, and the like.

TANGO 331 comprises a leucine zipper region at about amino acid residue 94 to about amino acid residue 115 (i.e., 94 LeaqeehLeawwlqLkseypdL 115). Leucine zipper regions are known to be involved in dimerization of proteins. Leucine zipper regions interact with one another, leading to formation of homo- or hetero-dimers between proteins, depending on their identity. The presence in TANGO 331 of a leucine zipper region is a further indication that this protein is involved in protein-protein interactions.

TANGO 331 shares amino acid and nucleic acid homology with a Chinese hamster protein designated HT, and thus is involved in corresponding physiological processes in humans. An alignment of the amino acid sequences of (human) TANGO 331 and Chinese hamster protein HT is shown in Figure 6E. In this alignment (made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4), the proteins are 71.9% identical. An alignment of the nucleotide sequences of the ORFs encoding (human) TANGO 331 and Chinese hamster protein HT is shown in Figures 6F through 6J. The two ORFs are 74.5% identical, as assessed using the same software and parameters.

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 331 protein includes an approximately 24 (i.e., 22, 23, 24, 25, or 26) amino acid residue signal peptide (amino acid residues 1 to 24 of SEQ ID NO: 56; SEQ ID NO: 57) preceding the mature TANGO 331 protein (i.e., approximately amino acid residues 25 to 353 of SEQ ID NO: 56; SEQ ID NO: 58). Mature human TANGO 331 is a secreted protein.

Figure 6D depicts a hydrophilicity plot of human TANGO 331 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region

which corresponds to amino acid residues 1 to 24 of SEQ ID NO: 56 is the signal sequence of human TANGO 331 (SEQ ID NO: 57). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 331 protein from about amino acid residue 140 to about amino acid residue 170 appears to be located at or near the surface of the protein, while the region from about amino acid residue 115 to about amino acid residue 130 appears not to be located at or near the surface.

The predicted molecular weight of human TANGO 331 protein without modification and prior to cleavage of the signal sequence is about 38.2 kilodaltons. The predicted molecular weight of the mature human TANGO 331 protein without modification and after cleavage of the signal sequence is about 35.6 kilodaltons.

Tissue distribution of TANGO 331 mRNA was determined by Northern blot hybridization. Northern blot hybridizations with the various RNA samples were performed using standard Northern blotting conditions and washing under stringent conditions (i.e., 0.2× SSC at 65°C). The DNA probe used in the Northern Blot experiments was radioactively labeled with 32P-dCTP using the PRIME-IT™ kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters having human mRNA disposed thereon (MULTITISSUE™ Northern I and MULTITISSUE™ Northern II obtained from Clontech, Palo Alto, CA) were probed in EXPRESSHYB™ hybridization solution (Clontech) and washed at high stringency according to the manufacturer's recommendations.

Two isoforms of human TANGO 331 were identified using this Northern blot analysis, indicating that TANGO 331 can have a splice varient. One isoform (corresponding to the larger message) can be a transmembrane protein (frizzled-like) and the other (i.e., smaller) isoform can be a secreted form. The two isoforms exhibit a clear pattern of tissue specificity. On the multiple tissue blot from Clonetech, the large transcript is found in almost all tissues, whereas the smaller message is expressed mainly in heart, skeletal muscle, placenta, and pancreas tissues.

TANGO 331 can be expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in E. coli and the fusion polypeptide is isolated and characterized. Specifically, TANGO 331 can be fused with GST and this fusion polypeptide can expressed in *E. coli*, e.g., in strain PEB199. Expression of the GST-

5   TANGO 331 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide can be purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography, e.g., using glutathione-substituted beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide can be

10   determined.

To express the TANGO 331 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) can be used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation

15   site. A DNA fragment encoding the entire TANGO 331 protein and an HA tag (Wilson et al. (1984) Cell 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment can be cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the TANGO 331 DNA sequence is amplified

20   by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the TANGO 331 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the TANGO 331 coding sequence. The PCR

25   amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the TANGO 331 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (e.g., one or more of strains HB101,

DH5a, SURE, available from Stratagene Cloning Systems, La Jolla, CA), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

5   COS cells are subsequently transfected using the TANGO 331-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods of transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory

10   Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the TANGO 331 polypeptide can be detected by radiolabelling (^{35}S-methionine or ^{35}S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring

15   Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labelled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 millimolar NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 millimolar Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody.

20   Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the TANGO 331 coding sequence can be cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the TANGO 331 polypeptide can be detected by

25   radiolabelling and immunoprecipitation using an TANGO 331 specific monoclonal antibody.

The human TANGO 331 gene was mapped using the Genebridge 4 Human Radiation hybrid mapping panel with ATTATTCAGAAGGATGTCCCGTGG (SEQ ID NO: 99) as the forward primer and CCTCCTGATTACCTACAATGGTC

(SEQ ID NO: 100) as the reverse primer. The human TANGO 331 gene maps to human 22q11-q13. Flanking markers for this region are WI-4572 and WI-8917. The schizophrenia 4 (sczd4) locus also maps to this region of the human chromosome. Also mapping to this region of the human chromosome are the following genes:

5   transcription factor 20 (tcf20), Benzodiazepine receptor, peripheral type (bzrp), Arylsulfatase A (arsa), diaphorase (NADH); cytochrome b-5 reductase (dia1), and Solute carrier family 5 (sodium/glucose transporter), member 1 (slca1). This region is syntenic to mouse chromosome 15. The stargazer (stg), gray tremor (gt), brachyury modifier 2 (Brm2), bronchial hyperresponsiveness 2 (Bhr2), loss of righting induced by

10   ethanol 5 (Lore5), fluctuating asymmetry QTL 8 (Faq8), jerky (Jrk), belted (bt), and koala (Koa) loci also map to this region of the mouse chromosome, several of which are neuromuscular related.

       <u>Biological function of TANGO 331 proteins, nucleic acids encoding them, and</u>
15   <u>modulators of these molecules</u>

              TANGO 331 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that TANGO 331 is expressed in human mammary epithelial tissue and human heart tissue, TANGO 331 protein is involved in

20   one or more biological processes which occur in mammary epithelial tissue, in other epithelial tissues, and in heart tissue. In particular, TANGO 331 is involved in modulating growth, proliferation, survival, differentiation, and activity of cells including, but not limited to, epithelial cells (e.g., mammary epithelial cells) of the animal in which it is normally expressed. Thus, TANGO 331 has a role in disorders

25   which affect these cells and their growth, proliferation, survival, differentiation, and activity. TANGO 331 is therefore involved in physiological processes such as maintenance of epithelia, carcinogenesis, modulation and storage of protein factors and metals, and lactation. Furthermore, because TANGO 331 is expressed in human mammary epithelial cells, it also has a role in nutrition of human infants (e.g.,

providing nutrients such as minerals to infants and providing protein factors not synthesized by infants) and in disorders which affect them. Thus, TANGO 331 is involved in a number of disorders such as breast cancer, insufficient lactation, infant nutritional and growth disorders, and the like. TANGO 331 polypeptides, nucleic

5   acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

Because TANGO 331 exhibits expression in the heart, TANGO 331 nucleic acids, proteins, and modulators thereof can be used to treat heart disorders. Examples of heart disorders with which TANGO 331 can be involved include ischemic

10   heart disease, atherosclerosis, hypertension, angina pectoris, hypertrophic cardiomyopathy, and congenital heart disease. TANGO 331 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

In another example, TANGO 331 polypeptides, nucleic acids, and

15   modulators thereof, can be involved in normal and aberrant functioning of skeletal muscle tissue, and can thus be involved in disorders of such tissue. Examples of skeletal muscle disorders include muscular dystrophy (e.g., Duchenne muscular dystrophy, Becker muscular dystrophy, Emery-Dreifuss muscular dystrophy, limb-girdle muscular dystrophy, facioscapulohumeral muscular dystrophy, myotonic

20   dystrophy, oculopharyngeal muscular dystrophy, distal muscular dystrophy, and congenital muscular dystrophy), motor neuron diseases (e.g., amyotrophic lateral sclerosis, infantile progressive spinal muscular atrophy, intermediate spinal muscular atrophy, spinal bulbar muscular atrophy, and adult spinal muscular atrophy), myopathies (e.g., inflammatory myopathies (e.g., dermatomyositis and polymyositis),

25   myotonia congenita, paramyotonia congenita, central core disease, nemaline myopathy, myotubular myopathy, and periodic paralysis), and metabolic diseases of muscle (e.g., phosphorylase deficiency, acid maltase deficiency, phosphofructokinase deficiency, debrancher enzyme deficiency, mitochondrial myopathy, carnitine deficiency, carnitine palmityl transferase deficiency, phosphoglycerate kinase deficiency, phosphoglycerate

mutase deficiency, lactate dehydrogenase deficiency, and myoadenylate deaminase deficiency). TANGO 331 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

5   In another example, TANGO 331 polypeptides, nucleic acids, and modulators thereof can be used to treat placental disorders, such as toxemia of pregnancy (e.g., preeclampsia and eclampsia), placentitis, and spontaneous abortion. TANGO 331 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

10   In another example, TANGO 331 polypeptides, nucleic acids, and modulators thereof can be used to treat placental disorders, such as toxemia of pregnancy (e.g., preeclampsia and eclampsia), placentitis, and spontaneous abortion.

          Presence in TANGO 331 of numerous EGF-like domains, including the laminin-like EGF-like domain indicates that TANGO 331 is involved in extracellular

15 binding of proteins, including both other secreted proteins (e.g., growth factors and cytokines) and cell-surface proteins. Binding of TANGO 331 to other secreted proteins modulates their activity, their rate of uptake by cells, and their rate of degradation. Binding of TANGO 331 to cell surface proteins modulates their activity, including, for example, their ability to bind with other secreted proteins, and transmits a signal to the

20 cell expressing the cell-surface protein. Presence in TANGO 331 of a TNFR/NGFR cysteine-rich region domain is further indicative of the ability of TANGO 331 to bind with growth factors and cytokines. Thus, TANGO 331 is involved in a number of proliferative and immune disorders including, but not limited to, cancers (e.g., breast cancer), autoimmune disorders, insufficient or inappropriate host responses to

25 infection, acquired immune deficiency syndrome, and the like. TANGO 331 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

          The fact that TANGO 331 has a metallothionein-like region is indicative of the ability of TANGO 331 to bind with metal ions, including nutritionally required

metal ions (e.g., calcium, magnesium, zinc, manganese, cobalt, iron, and the like). Thus, TANGO 331 is involved in binding with essential minerals and in delivering them to their proper body locations. TANGO 331 is also involved in binding excess or toxic metal ions so that they can be excreted. TANGO 331 is thus involved in

5 disorders involving insufficient or inappropriate localization of metal ions. Such disorders include, but are not limited to, malnutrition and mineral deficiency disorders, hemochromatosis, inappropriate calcification of body tissues, bone disorders such as osteoporosis, and the like. TANGO 331 polypeptides, nucleic acids, or modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more

10 of these disorders.

Mapping of the human TANGO 331 gene to chromosomal region 22q11-q13 is an indication of disorders with which its expression (or non- or aberrant-expression) can be associated. For example, arylsulfatase A is associated with Metachromatic leukodystrophy. Diaphorase (NADH:cytochrome b-5 reductase) is

15 associated with methemoglobinemia, types I and II. Solute carrier family 5 (sodium/glucose transporter), member 1 is associated with glucose/galactose malabsorption. The gene designated schizophrenia 4 is associated with schizophrenia and velocardiofacial syndrome, as described in Online Mendelian Inheritance in Man, Johns Hopkins University, Baltimore, MD. MIM Number: 600850:12/7/98. (World

20 Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/). These mapping data indicate that TANGO 331 polypeptides, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, inhibit, prevent, or alleviate one or more of these disorders.

TANGO 332

25 A cDNA clone (designated jlhbab463g12) encoding at least a portion of human TANGO 332 protein was isolated from a human adult brain cDNA library. The human TANGO 332 protein is predicted by structural analysis to be a secreted protein.

The full length of the cDNA encoding human TANGO 332 protein (Figure 7; SEQ ID NO: 59) is 2730 nucleotide residues. The ORF of this cDNA,

nucleotide residues 173 to 2185 of SEQ ID NO: 59 (i.e., SEQ ID NO: 60), encodes a 671-amino acid transmembrane protein (Figure 7; SEQ ID NO: 61).

The invention thus includes purified human TANGO 332 protein, both in the form of the immature 671 amino acid residue protein (SEQ ID NO: 61) and in the form of the mature, approximately 649 amino acid residue protein (SEQ ID NO: 63). Mature human TANGO 332 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature TANGO 332 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature human TANGO 332 proteins, the invention includes fragments, derivatives, and variants of these TANGO 332 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as TANGO 332 polypeptides of the invention or TANGO 332 proteins of the invention.

The invention also includes nucleic acid molecules which encode a TANGO 332 polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 59 or some portion thereof, such as the portion which encodes mature TANGO 332 protein, immature TANGO 332 protein, or a domain of TANGO 332 protein. These nucleic acids are collectively referred to as TANGO 332 nucleic acids of the invention.

TANGO 332 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features, as indicated by the conservation of amino acid sequence between human TANGO 332 protein, human brain-enriched hyaluronan-binding factor (BEF), as shown in Figures 7G and 7H, and murine brevican protein, as shown in Figures 7I to 7K. This conservation is further indicated by conservation of nucleotide sequence between the ORFs encoding human TANGO 332 protein and murine brevican protein, as shown in Figures 7L through 7U.

A common domain present in TANGO 332 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10

amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino

5   acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 332 protein contains a signal sequence corresponding to about amino acid residues 1 to 22 of SEQ ID NO: 61 (SEQ

10   ID NO: 62). The signal sequence is cleaved during processing of the mature protein.

TANGO 332 proteins are secreted proteins. The mature form of human TANGO 332 protein has the amino acid sequence of approximately amino acid residues 23 to 671 of SEQ ID NO: 61.

TANGO 332 proteins typically comprise a variety of potential post-

15   translational modification sites (often within an extracellular domain), such as those described herein in Table IX, as predicted by computerized sequence analysis of TANGO 332 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 332 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In

20   certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table IX.

Table IX

Type of Potential Modification Site or Domain	Amino Acid Residues of SEQ ID NO: 61	Amino Acid Sequence
N-glycosylation site	130 to 133	NDSG
	337 to 340	NQTG
Protein kinase C phosphorylation site	67 to 69	SRR
	74 to 76	SPR
	165 to 167	SAR
	212 to 214	TVR
	219 to 221	TPR
	310 to 312	SVR
	319 to 321	SQR
	545 to 547	TPR
	615 to 617	SGR
Casein kinase II phosphorylation site	29 to 32	SSED
	116 to 119	SLTD
	219 to 222	TPRE
	269 to 272	TLEE
	382 to 385	TVTE
	386 to 389	TLEE
	397 to 400	TESE
	419 to 422	STPE
	430 to 433	TLLE
	446 to 449	SEEE
	545 to 548	TPRE
	558 to 561	TLVE
Tyrosine kinase phosphorylation site	128 to 135	RPNDSGIY
	451 to 459	KALEEEEKY

Table IX (Continued)

N-myristoylation site	47 to 52	GVLGGA
	133 to 138	GIYRCE
	142 to 147	GIDDSS
	174 to 179	GAQEAC
	183 to 188	GAHIAT
	281 to 286	GAEIAT
	288 to 293	GQLYAA
	297 to 302	GLDHCS
	324 to 329	GGLPGV
	403 to 408	GAIYSI
	414 to 419	GGGGSS
	576 to 581	GVPRGE
	586 to 591	GSSEGA
Immunoglobulin-/major histocompatibility protein-like (Ig-/MHC-like) domain	50 to 141	See Fig. 7
Extracellular link domain	156 to 251	See Fig. 7
	257 to 353	See Fig. 7

Among the domains that occur in TANGO 332 protein are an Ig-/MHC-like domain and a pair of extracellular link domains. In one embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to one of these domains. In other embodiments, the protein has at least one Ig-/MHC-like domain and one extracellular link domain described herein in Table IX. In other embodiments, the protein has at least one Ig-/MHC-like domain and at least two extracellular link domains.

Ig-/MHC-like domains are conserved among immunoglobulin (Ig) constant (CL) regions and one of the three extracellular domains of major histocompatibility proteins (MHC). The presence in TANGO 332 of an Ig-/MHC-like domain indicates that the corresponding region of TANGO 332 is structurally similar to this conserved extracellular region.

Extracellular link domains occur in hyaluronan- (HA-)binding proteins. Proteins having this domain include cartilage link protein, proteoglycans such as aggrecan, brevican, neurocan, and versican, CD44 antigen (the primary cell surface receptor for HA), and tumor necrosis factor-inducible protein TSG-6. Presence of a pair of extracellular link domains in TANGO 332 indicates that this protein is also involved in HA-binding, and therefore is involved in physiological processes such as cartilage (and other tissue) organization, extracellular matrix organization, neural growth and branching, and cell-to-cell and cell-to-matrix interactions. Involvement of TANGO 332 in these processes implicates this protein in disorders such as tumor growth and metastasis, movement of cells (e.g., leukocytes) through extracellular matrix, inappropriate inflammation, and the like.

Brevican is a murine nervous system-specific chondroitin sulfate proteoglycan which binds in a calcium-dependent manner with two classes of sulfated glycolipids, namely sulfatides and HNK-1-reactive sulfoglucuronylglycolipids (Miura et al. (1999) *J. Biol. Chem.* 274:11431-11438). A human orthologue, designated BEF ('Brain-Enriched hyaluronan-binding Factor'), of murine brevican is expressed by human glioma cells, but not by brain tumors of non-glial origin (P.C.T. application publication number WO98/31800; Zhang et al. (1998) *J. Neurosci.* 18:2370-2376). Those authors suggested that cleavage of that human orthologue mediates glioma cell invasion *in vivo.*

An alignment of the amino acid sequences of TANGO 332 and BEF protein is shown in Figures 7G and 7H. In this alignment (made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4), the proteins are 75.7% identical,

although it is seen that TANGO 332 includes two domains (one from about amino acid residue 152 to about residue 208, and the other near the carboxyl terminus of TANGO 332) which do not occur in BEF protein. It is likely that these two regions account for the differences between the physiological roles of TANGO 332 and BEF.

5　　　　An alignment of the amino acid sequences of (human) TANGO 332 and murine brevican protein is shown in Figures 7I through 7K. In this alignment (made using the ALIGN software {Myers and Miller (1989) *CABIOS*, ver. 2.0}; pam120.mat scoring matrix; gap opening pentaly = 12, gap extension penalty = 4), the proteins are 75.5% identical, although it is seen that murine brevican protein includes a domain

10　which does not occur in TANGO 332 protein, this domain is present from about amino acid residue 626 to the carboxyl terminus of murine brevican protein. An alignment of the nucleotide sequences of the ORFs encoding (human) TANGO 332 and murine brevican protein is shown in Figures 7L through 7U. The two ORFs are 62.6% identical, as assessed using the same software and parameters.

15　　　　TANGO 332 exhibits many of the same properties as BEF. TANGO 332 is also related to murine brevican protein, and thus is involved with corresponding physiological processes (i.e., such as those described above) in humans. For example, TANGO 332 modulates intracellular binding and migration of cells in a tissue or extracellular matrix. However, the absence from BEF of one of the two extracellular

20　link domains present in TANGO 332 indicates that one or more of the subunit structure, the tissue specificity, and the binding specificity of TANGO 332 and BEF proteins differ. Thus, TANGO 332 is involved in many of the physiological processes and disorders in which BEF protein is involved. Like murine brevican and other proteoglycans, TANGO 332 acts *in vivo* as a tissue organizing protein, influences

25　growth and maturation of tissues in which it is expressed, modulates growth factor-mediated activities, modulates structural features of tissues (e.g., collagen fibrillogenesis), modulates tumor cell growth and invasivity, and influences neurite growth and branching.

- 100 -

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 332 protein includes an approximately 22 (i.e., 20, 21, 22, 23, or 24) amino acid residue signal peptide (amino acid residues 1 to 22 of SEQ ID NO: 61; SEQ ID NO: 62) preceding the mature

5   TANGO 332 protein (i.e., approximately amino acid residues 23 to 671 of SEQ ID NO: 61; SEQ ID NO: 63). Human TANGO 332 protein is a secreted protein, as assessed using the secretion assay described herein. Secreted TANGO 332 proteins having approximate sizes of 148 kilodaltons and 100 kilodaltons could be detected using this assay.

10   Figure 7F depicts a hydrophilicity plot of human TANGO 332 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 22 of SEQ ID NO: 61 is the signal sequence of human TANGO 332 (SEQ ID NO: 62). As described elsewhere herein,

15   relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 332 protein from about amino acid residue 445 to about amino acid residue 475 appears to be located at or near the surface of the protein, while the region from about amino acid residue 45 to

20   about amino acid residue 62 appears not to be located at or near the surface.

            The predicted molecular weight of human TANGO 332 protein without modification and prior to cleavage of the signal sequence is about 71.7 kilodaltons. The predicted molecular weight of the mature human TANGO 332 protein without modification and after cleavage of the signal sequence is about 69.5 kilodaltons.

25

            Biological function of TANGO 332 proteins, nucleic acids encoding them, and modulators of these molecules

            TANGO 332 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not

expressed. Based on the observation that TANGO 332 is expressed in human adult brain tissue, TANGO 332 protein is involved in one or more biological processes which occur in these tissues. In particular, TANGO 332 is involved in modulating growth, proliferation, survival, differentiation, and activity of cells including, but not

5   limited to, adult brain cells of the animal in which it is normally expressed. Thus, TANGO 332 has a role in disorders which affect these cells and their growth, proliferation, survival, differentiation, interaction, and activity. Examples of such disorders include, by way of example and not limitation, disorders of neural connection establishment or maintenance, impaired cognitive function, dementia, senility,

10   Alzheimer's disease, mental retardation, brain tumors (e.g., gliomas such as astrocytomas, endophytic and exophytic retinoblastomas, ependymomas, gangliogliomas, mixed gliomas, nasal gliomas, optic gliomas, and Schwannomas, and other brain cell tumors such as medulloblastomas, pituitary adenomas, teratomas, etc.), and the like.

15   Homology of human TANGO 332 with murine brevican protein and with human brevican homolog BEF indicates that TANGO 332 has physiological functions in humans analogous to the functions of these proteins. Brevican is a member of the aggrecan/versican family of proteoglycans, and has a hyaluronic acid-binding domain in its amino terminal region and a lectin-like domain in its carboxyl

20   terminal region. Expression of brevican is highly specific to brain tissue, and increases as the mammalian brain develops. Thus, brevican is involved in maintaining the extracellular environment of mature brain tissue and is a constituent of adult brain extracellular matrix. TANGO 332 is involved in modulating cell-to-cell adhesion, tissue and extracellular matrix invasivity of cells, and the like. Thus, TANGO 332 is

25   involved in disorders in which these physiological processes are relevant. Such disorders include, for example, loss of control of cell growth, tumor metastasis, malformation of neurological connections, inflammation, immune and autoimmune responses, and the like.

In addition, presence in TANGO 332 of extracellular link domains indicates that this protein is involved in physiological processes involving structure and function of extracellular matrices and interaction of cells with such matrices and with each other. This is further evidence that TANGO 332 is involved in disorders such as inappropriate inflammation, tumor metastasis, inappropriate leukocyte extravasation, localization, and reactivity, and the like.

TANGO 332-related molecules can be used to modulate one or more of the activities in which TANGO 332 is involved and can also be used to prevent, diagnose, or treat one or more of the disorders in which TANGO 332 is involved.

Tables A and B summarize sequence data corresponding to the human proteins herein designated INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, and TANGO 332.

Table A

Protein Designation	SEQ ID NOs			Depicted in Figure #	ATCC® Accession #
	cDNA	ORF	Protein		
INTERCEPT 217	1	2	3	1	PTA-147
INTERCEPT 297	9	10	11	2	PTA-147
TANGO 276	33	34	35	3	PTA-150
TANGO 292	38	39	40	4	207230
TANGO 325	46	47	48	5	PTA-147
TANGO 331	54	55	56	6	PTA-147
TANGO 332	59	60	61	7	PTA-151

Table B

Protein Desig.	Signal Sequence	Mature Protein	Extracellular Domain(s)	Transmembrane Domain(s)	Cytoplasmic Domain(s)
			SEQ ID NOs		
INTERCEPT 217	4	5	6	7	8
INTERCEPT 297	(12)	13	14, 15, 16, 17, 18	(12), 19, 20, 21, 22, 23, 24, 25, 26, 27	28, 29, 30, 31, 32
TANGO 276	36	37	37	N/A	
TANGO 292	41	42	43	44	45
TANGO 325	49	50	51	52	53
TANGO 331	57	58	58	N/A	N/A
TANGO 332	62	63	63	N/A	N/A
			Amino Acid Residues		
INTERCEPT 217	1-20	21-455	21-383	384-403	404-455
INTERCEPT 297	(1-18)	19-371	19-47, 110-118, 162-175, 234-260, 313-319	(1-18), 48-68, 89-109, 119-137, 145-161, 176-192, 216-233, 261-283, 293-312, 320-336	69-88, 138-144, 193-215, 284-292, 337-371
TANGO 276	1-20	21-243	21-243	N/A	N/A
TANGO 292	1-17	18-226	18-113	114-138	139-226
TANGO 325	1-31	32-622	32-529	530-547	548-622
TANGO 331	1-24	25-353	25-353	N/A	N/A
TANGO 332	1-22	23-671	23-671	N/A	N/A

Various aspects of the invention are described in further detail in the following subsections.

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a biologically active portion thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein-encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kilobases, 4 kilobases, 3 kilobases, 2 kilobases, 1 kilobases, 0.5 kilobases, or 0.1 kilobases of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of all or a portion of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or a complement thereof, or which has a nucleotide sequence comprising one of these sequences, can be isolated using

5   standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92 as a hybridization probe, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook et al., eds., *Molecular Cloning: A*

10   *Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can

15   be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of

20   the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby

25   forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding a full length polypeptide of the invention for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a polypeptide of the invention. The nucleotide sequence

- 107 -

determined from the cloning one gene allows for the generation of probes and primers designed for use in identifying and/or cloning homologs in other cell types, e.g., from other tissues, as well as homologs from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically

5   comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 15, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, or 400 or more consecutive nucleotides of the sense or anti-sense sequence of one of any of SEQ ID NOs: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92, or of a naturally occurring mutant of one of SEQ ID NO: 1, 2, 9,

10   10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, and 92.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences encoding the same protein molecule encoded by a selected nucleic acid molecule. The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme,

15   or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, e.g., detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

20   A nucleic acid fragment encoding a biologically active portion of a polypeptide of the invention can be prepared by isolating a portion of one of SEQ ID NO: 2, 10, 34, 39, 47, 55, 60, 82, and 92, expressing the encoded portion of the polypeptide protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the polypeptide.

25   The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence of SEQ ID NO: 2, 10, 34, 39, 47, 55, 60, 82, or 92.

- 108 -

In addition to the nucleotide sequences of SEQ ID NOs: 2, 10, 34, 39, 47, 55, 60, 82, and 92, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence can exist within a population (e.g., the human population). Such genetic polymorphisms can exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus.

As used herein, the phrase "allelic variant" refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence.

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding proteins of the invention from other species (homologs), which have a nucleotide sequence which differs from that of the specific proteins described herein are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologs of a cDNA of the invention can be isolated based on their identity to human nucleic acid molecules using the cDNAs described herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble form of a membrane-bound protein of the invention isolated based on its hybridization to a nucleic acid molecule encoding all or part of the membrane-bound form. Likewise, a

cDNA encoding a membrane-bound form can be isolated based on its hybridization to a nucleic acid molecule encoding all or part of the soluble form.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 (25, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 4500, or 4928) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or a complement thereof. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or a complement thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention sequence that can exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type

- 110 -

sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologs of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologs of various species (e.g., murine and human) may be essential for activity and thus would not be likely targets for alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 40% identical, 50%, 60%, 70%, 80%, 90%, 95%, or 98% identical to the amino acid sequence of one of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, such that one or more amino acid residue substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine,

- 111 -

tyrosine, cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Alternatively, mutations can be introduced randomly along all

5   or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant polypeptide that is a variant of a

10   polypeptide of the invention can be assayed for: (1) the ability to form protein:protein interactions with the polypeptide of the invention; (2) the ability to bind a ligand of the polypeptide of the invention (e.g., another protein identified herein); (3) the ability to bind to a modulator or substrate of the polypeptide of the invention; or (4) the ability to modulate a physiological activity of the protein, such as one of those disclosed herein.

15   The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a polypeptide of the invention, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense

20   nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' un-translated regions") are the 5' and 3' sequences

25   which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an

antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and

5   acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine,

10   N_6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine,

15   pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been sub-cloned in an

20   antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to

25   cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix.

- 113 -

An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach (1988) *Nature* 334:585-591) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of the invention can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent

- 114 -

No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (e.g., the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al. (1996) *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or anti-gene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup

- 115 -

(1996), *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by

5 the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which can combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and

10 specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996), *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra,* and Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support

15 using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al. (1989) *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a step-wise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment

20 (Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide can include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents

25 facilitating transport across the cell membrane (*see, e.g.,* Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810) or the blood-brain barrier (*see, e.g.,* PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.,* Krol et al. (1988)

Bio/Techniques 6:958-976) or intercalating agents (*see, e.g.,* Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

5

II. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide of the invention. In one embodiment, the native polypeptide can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, polypeptides of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide of the invention can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such

preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein (e.g., the amino acid sequence shown in any of SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98), which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Preferred polypeptides have the amino acid sequence of one of SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98. Other useful proteins are substantially identical (e.g., at least about 40%, preferably 50%, 60%, 70%, 80%, 90%, 95%, or 99%) to any of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, and retain the functional activity of the protein of the corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two

sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions (e.g., overlapping positions) × 100). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules. *Id.* When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably biologically active) of a polypeptide of the invention operably linked to a heterologous polypeptide (i.e., a polypeptide other than the same polypeptide of the invention). Within the

5 fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the amino-terminus or the carboxyl-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the

10 polypeptide of the invention is fused to the carboxyl terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal sequence at its amino terminus. For example, the native signal sequence of a

15 polypeptide of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human

20 placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin

25 fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell

(receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction can be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (e.g.,

5   promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by

10   standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric

15   gene sequence (*see, e.g.,* Ausubel et al., *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

20   A signal sequence of a polypeptide of the invention (e.g., the signal sequence in one of SEQ ID NO: 3, 4, 11, 12, 35, 36, 40, 41, 48, 49, 56, 57, 61, 62, 83, 84, 93, and 94) can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during

25   secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal sequence, as well as to the signal sequence itself and to the polypeptide in the absence of the signal sequence (i.e., the cleavage products). In

- 121 -

one embodiment, a nucleic acid sequence encoding a signal sequence of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the

5   expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

10   In another embodiment, the signal sequences of the present invention can be used to identify regulatory sequences, e.g., promoters, enhancers, repressors. Since signal sequences are the most amino-terminal sequences of a peptide, the nucleic acids which flank the signal sequence on its amino-terminal side are likely regulatory sequences which affect transcription. Thus, a nucleotide sequence which encodes all or

15   a portion of a signal sequence can be used as a probe to identify and isolate signal sequences and their flanking regions, and these flanking regions can be studied to identify regulatory elements therein.

             The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as

20   either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, e.g., discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a

25   downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side

effects in a subject relative to treatment with the naturally occurring form of the protein.

Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the protein of the invention for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (*see, e.g.,* Narang (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, re-naturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes amino terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene

5   libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional

10   mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

An isolated polypeptide of the invention, or a fragment thereof, can be

15   used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30 or more) amino acid residues of the amino acid sequence

20   of one of SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, e.g., hydrophilic regions. Figures 1F, 1M, 2D,

25   3E, 4E, 4M, 5F, 6D, and 7F are hydrophobicity plots of the proteins of the invention. These plots or similar analyses can be used to identify hydrophilic regions.

An immunogen typically is used to prepare antibodies by immunizing a suitable (i.e., immunocompetent) subject such as a rabbit, goat, mouse, or other mammal or vertebrate. An appropriate immunogenic preparation can contain, for

- 124 -

example, recombinantly-expressed or chemically-synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or a similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The terms "antibody" and "antibody substance" as used interchangeably herein refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, e.g., a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')$_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. Preferred polyclonal antibody compositions are ones that have been selected for antibodies directed against (i.e., which bind specifically with) one or more polypeptides of the invention. Particularly preferred polyclonal antibody preparations are ones that contain only antibodies directed against one or more polypeptides of the invention. Particularly preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen compositions made using a non-human host cell for recombinant expression of a polypeptide of the invention. In such a manner, the only human epitope or epitopes recognized by the resulting antibody compositions

- 125 -

raised against this immunogen will be present as part of a polypeptide or polypeptides of the invention.

The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be harvested or isolated from the subject (e.g., from the blood or serum of the subject) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. Alternatively, antibodies which bind specifically with a protein or polypeptide of the invention can be selected (e.g., partially purified) or purified using chromatographic methods, such as affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention can be produced as described herein, and covalently or non-covalently coupled with a solid support such as, for example, a chromatography column. The column thus exhibits specific affinity for antibody substances which bind specifically with the protein of the invention, and these antibody substances can be purified from a sample containing antibody substances directed against a large number of different epitopes, thereby generating a substantially purified antibody substance composition, i.e., one that is substantially free of antibody substances which do not bind specifically with the protein. By a substantially purified antibody composition is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than those on the desired protein or polypeptide of the invention, preferably at most 20%, more preferably at most 10%, most preferably at most 5% (by dry weight), of the sample is contaminating antibodies. A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein or polypeptide of the invention.

At an appropriate time after immunization, e.g., when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature*
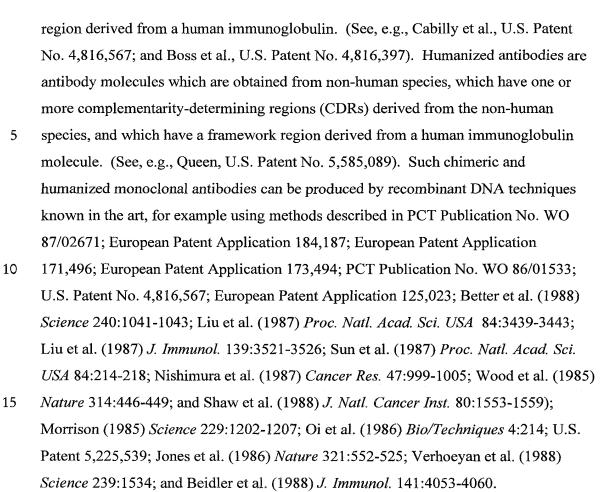
256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (*see generally Current Protocols in Immunology* (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SURFZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J.* 12:725-734.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. A chimeric antibody is a molecule in which different portions of the antibody amino acid sequence are derived from different animal species, such as those having a variable region derived from a murine monoclonal antibody and a constant

region derived from a human immunoglobulin. (See, e.g., Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816,397). Humanized antibodies are antibody molecules which are obtained from non-human species, which have one or more complementarity-determining regions (CDRs) derived from the non-human

5   species, and which have a framework region derived from a human immunoglobulin molecule. (See, e.g., Queen, U.S. Patent No. 5,585,089). Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application

10   171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al. (1987) *Cancer Res.* 47:999-1005; Wood et al. (1985)

15   *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison (1985) *Science* 229:1202-1207; Oi et al. (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

         Completely human antibodies are particularly desirable for therapeutic

20   treatment of human patients. Such antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against

25   the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies,

see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, *see, e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent

5   5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a

10   selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope (Jespers et al., 1994, *Bio/technology* 12:899-903).

An antibody directed against a polypeptide of the invention (e.g., monoclonal antibody) can be used to isolate the polypeptide by standard techniques,

15   such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment

20   regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable

25   prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of
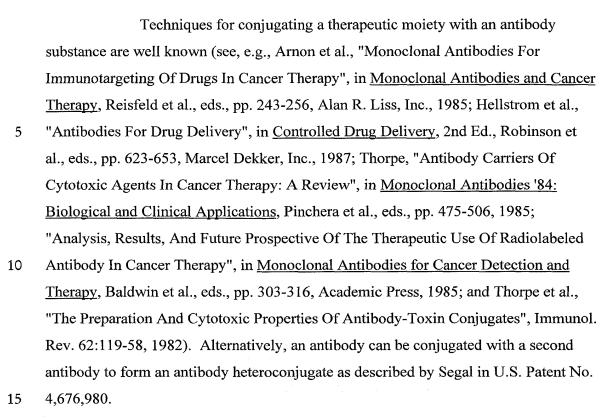
- 129 -

bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^3H.

Further, an antibody substance can be conjugated with a therapeutic moiety such as a cytotoxin, a therapeutic agent, or a radioactive metal ion. Cytotoxins and cytotoxic agents include any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, puromycin, and analogs or homologs of these compounds. Therapeutic agents include, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil, and decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine {BSNU}, lomustine {CCNU}, cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin {formerly daunomycin} and doxorubicin), antibiotics (e.g., dactinomycin {formerly actinomycin}, bleomycin, mithramycin, and anthramycin {AMC}), and anti-mitotic agents (e.g., vincristine and vinblastine).

The conjugates of the invention can be used to modify a biological response; the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety can be a protein or polypeptide which exhibits a desired biological activity. Such proteins include, for example, toxins such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; proteins such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; and biological response modifiers such as lymphokines, interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), and other growth factors.
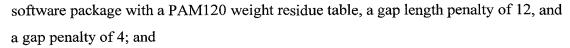
Techniques for conjugating a therapeutic moiety with an antibody substance are well known (see, e.g., Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies and Cancer Therapy, Reisfeld et al., eds., pp. 243-256, Alan R. Liss, Inc., 1985; Hellstrom et al.,

5   "Antibodies For Drug Delivery", in Controlled Drug Delivery, 2nd Ed., Robinson et al., eds., pp. 623-653, Marcel Dekker, Inc., 1987; Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological and Clinical Applications, Pinchera et al., eds., pp. 475-506, 1985; "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled

10   Antibody In Cancer Therapy", in Monoclonal Antibodies for Cancer Detection and Therapy, Baldwin et al., eds., pp. 303-316, Academic Press, 1985; and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev. 62:119-58, 1982). Alternatively, an antibody can be conjugated with a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No.

15   4,676,980.

Accordingly, in one aspect, the invention provides substantially purified antibodies or fragment thereof, and non-human antibodies or fragments thereof, which antibodies or fragments specifically bind with a polypeptide having an amino acid sequence which comprises a sequence selected from the group consisting of

20   (i) SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98;

(ii) the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151;

(iii) a fragment of at least 15 amino acid residues of the amino acid sequence of

25   SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98;

(iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98, wherein the percent identity is determined using the ALIGN program of the GCG

software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and

(v) an amino acid sequence which is encoded by a nucleic acid molecule, the complement of which hybridizes with a nucleic acid molecule having the sequence of

5   SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or with a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, under conditions of hybridization of 6× SSC (standard saline citrate buffer) at 45°C and washing in 0.2× SSC, 0.1% SDS at 65°C.

In another aspect, the invention provides non-human antibodies or

10   fragments thereof, which antibodies or fragments specifically bind with a polypeptide having an amino acid sequence which comprises a sequence selected from the group consisting of:

(i) SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98;

15   (ii) the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151;

(iii) a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98;

(iv) an amino acid sequence which is at least 95% identical to the amino acid

20   sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and

(v) an amino acid sequence which is encoded by a nucleic acid molecule, the

25   complement of which hybridizes with a nucleic acid molecule having the sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or with a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, under conditions of hybridization of 6× SSC (standard saline citrate buffer) at 45°C and washing in 0.2× SSC, 0.1% SDS at 65°C. Such non-human antibodies can be goat,

- 132 -

mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind with a polypeptide having an amino acid sequence which comprises a sequence selected from the group consisting of:

(i) SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98;

(ii) the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151;

(iii) a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98;

(iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and

(v) an amino acid sequence which is encoded by a nucleic acid molecule, the complement of which hybridizes with a nucleic acid molecule having the sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or with a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, under conditions of hybridization of 6× SSC (standard saline citrate buffer) at 45°C and washing in 0.2× SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

The substantially purified antibodies or fragments thereof can specifically bind with a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a cytoplasmic domain cytoplasmic membrane of a polypeptide of the

- 133 -

invention. In a particularly preferred embodiment, the substantially purified antibodies or fragments thereof, the non-human antibodies or fragments thereof, and/or the monoclonal antibodies or fragments thereof, of the invention specifically bind with a secreted sequence or with an extracellular domain of one of INTERCEPT 217,

5   INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, and TANGO 332. Preferably, the extracellular domain with which the antibody substance binds has an amino acid sequence selected from the group consisting of SEQ ID NOs: 6, 14-18, 37, 43, 51, 58, or 63.

Any of the antibody substances of the invention can be conjugated with

10   a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated with the antibody substances of the invention include an enzyme, a prosthetic group, a fluorescent material (i.e., a fluorophore), a luminescent material, a bioluminescent material, and a radioactive material (e.g., a radionuclide or a substituent comprising a radionuclide)..

15   The invention also provides a kit containing an antibody substance of the invention conjugated with a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody substance of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody

20   substance of the invention, a therapeutic moiety (preferably conjugated with the antibody substance), and a pharmaceutically acceptable carrier.

Still another aspect of the invention is a method of making an antibody that specifically recognizes one of INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, and TANGO 332. This method comprises

25   immunizing a vertebrate (e.g., a mammal such as a rabbit, goat, or pig) with a polypeptide. The polypeptide used as an immunogen has an amino acid sequence that comprises a sequence selected from the group consisting of:

(i) SEQ ID NOs: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, and 93-98;

- 134 -

(ii) the amino acid sequence encoded by a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151;

(iii) a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98;

5 (iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO: 3-8, 11-32, 35-37, 40-45, 48-53, 56-58, 61-63, 83-88, or 93-98, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and

10 (v) an amino acid sequence which is encoded by a nucleic acid molecule, the complement of which hybridizes with a nucleic acid molecule having the sequence of SEQ ID NO: 1, 2, 9, 10, 33, 34, 38, 39, 46, 47, 54, 55, 59, 60, 81, 82, or 92, or with a cDNA of a clone deposited as ATCC® PTA-147, PTA-150, 207230, or PTA-151, under conditions of hybridization of 6× SSC (standard saline citrate buffer) at 45°C and 15 washing in 0.2× SSC, 0.1% SDS at 65°C.

After immunization, a sample is collected from the vertebrate that contains an antibody that specifically recognizes the polypeptide with which the vertebrate was immunized. Preferably, the polypeptide is recombinantly produced using a non-human host cell. Optionally, an antibody substance can be further purified 20 from the sample using techniques well known to those of skill in the art. The method can further comprise making a monoclonal antibody-producing cell from a cell of the vertebrate. Optionally, antibodies can be collected from the antibody-producing cell.

III. Recombinant Expression Vectors and Host Cells

25 Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide of the invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into

which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal

5 mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA

10 techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell.

15 acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence

20 of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are

25 described, for example, in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by

- 136 -

those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or

5   peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (e.g., *E. coli*) or eukaryotic cells (e.g., insect cells (using baculovirus expression vectors), yeast cells or mammalian cells). Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the

10   recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a

15   protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is

20   introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL

25   (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al.,

Gene Expression Technology: Methods in Enzymology 185, Academic Press, San

Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on

host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene

expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion

5   promoter mediated by a co-expressed viral RNA polymerase (T7 gn1). This viral

polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident

lambda prophage harboring a T7 gn1 gene under the transcriptional control of the

lacUV 5 promoter.

        One strategy to maximize recombinant protein expression in *E. coli* is to

10   express the protein in a host bacteria with an impaired capacity to proteolytically cleave

the recombinant protein (Gottesman, *Gene Expression Technology: Methods in*

Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another

strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an

expression vector so that the individual codons for each amino acid are those

15   preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118).

Such alteration of nucleic acid sequences of the invention can be carried out by

standard DNA synthesis techniques.

        In another embodiment, the expression vector is a yeast expression

vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1

20   (Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell*

30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen

Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

        Alternatively, the expression vector is a baculovirus expression vector.

Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf

25   9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and

the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

        In yet another embodiment, a nucleic acid of the invention is expressed

in mammalian cells using a mammalian expression vector. Examples of mammalian

expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC

(Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both

5   prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al., *supra*.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of

10   suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific

15   promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example

20   the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory

25   sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral

- 139 -

promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid, or attenuated virus in which antisense nucleic acids are produced under the control of a

5 high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (*Reviews - Trends in Genetics*, Vol. 1(1) 1986).

Another aspect of the invention pertains to host cells into which a

10 recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not,

15 in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (e.g., *E. coli*) or eukaryotic cell (e.g., insect cells, yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via

20 conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host

25 cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., for resistance to

- 140 -

antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

In another embodiment, the expression characteristics of an endogenous nucleic acid within a cell, cell line, or microorganism (e.g., a INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332 nucleic acid, as described herein) can be modified by inserting a heterologous DNA regulatory element (i.e., one that is heterologous with respect to the endogenous gene) into the genome of the cell, stable cell line, or cloned microorganism. The inserted regulatory element can be operatively linked with the endogenous gene (e.g., INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332) and thereby control, modulate, or activate the endogenous gene. For example, an endogenous INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332 gene which is normally "transcriptionally silent" (i.e., a INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332 gene which is normally not expressed, or is normally expressed only at only a very low level) can be activated by inserting a regulatory element which is capable of promoting expression of the gene in the cell, cell line, or microorganism. Alternatively, a transcriptionally silent, endogenous INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332 gene can be activated by inserting a promiscuous regulatory element that works across cell types.

A heterologous regulatory element can be inserted into a stable cell line or cloned microorganism such that it is operatively linked with and activates expression of an endogenous INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332 gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art (described

e.g., in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide of the invention using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a sequences encoding a polypeptide of the invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a polypeptide of the invention have been introduced into their genome or homologous recombinant animals in which endogenous encoding a polypeptide of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide and for identifying and/or evaluating modulators of polypeptide activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the

endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing nucleic acid encoding a polypeptide of the invention (or a homologue thereof) into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986), and in Wakayama et al., 1999, Proc. Natl. Acad. Sci. USA 96:14984-14989. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide of the invention into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein

(e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (*see, e.g.,* Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (*see, e.g.,* Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (*see, e.g.,* Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication NOS. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, *see, e.g.,* Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991)

Science 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) *Nature* 385:810-813 and PCT Publication NOS. WO 97/07668 and WO 97/07669.

IV. Pharmaceutical Compositions

The nucleic acid molecules, polypeptides, and antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical

composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention and one or more additional active compounds.

The agent which modulates expression or activity can, for example, be a small molecule. For example, such small molecules include peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents and protein or polypeptide agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of these agents will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the agent to have upon the nucleic acid or polypeptide of the invention. Exemplary doses of a small molecule include milligram or microgram amounts per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram). Exemplary doses of a protein or polypeptide include gram, milligram or microgram amounts per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 5 grams per kilogram, about 100 micrograms per kilogram to about 500 milligrams per kilogram, or about 1 milligram per kilogram to about 50 milligrams per kilogram). It is furthermore understood that appropriate doses of one of these agents depend upon the

potency of the agent with respect to the expression or activity to be modulated. Such appropriate doses can be determined using the assays described herein. When one or more of these agents is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a

5   physician, veterinarian, or researcher can, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific agent employed, the age, body weight, general health, gender, and diet of the subject, the

10   time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral

15   (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens;

20   antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediamine-tetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose

25   vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic

water, Cremophor EL (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as

5   bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of

10   surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride in the composition. Prolonged absorption of the injectable compositions can

15   be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

        Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required,

20   followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium, and then incorporating the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a

25   powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

        Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with

excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches, and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl

acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes having one

5 or more monoclonal antibodies incorporated therein or thereon; e.g., liposomes comprising a monoclonal antibody which binds specifically with a virus antigen) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

10 It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association

15 with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

20 For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is

25 often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al. ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470), or by stereotactic injection (*see, e.g.,* Chen et al. (1994) *Proc. Natl. Acad. Sci. USA*

5   91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce

10   the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

15   The nucleic acid molecules, proteins, protein homologs, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (e.g., therapeutic

20   and prophylactic). For example, polypeptides of the invention can to used for all of the purposes identified herein in portions of the disclosure relating to individual types of protein of the invention (e.g., INTERCEPT 217 proteins, INTERCEPT 297 proteins, TANGO 276 proteins, TANGO 292 proteins, TANGO 325 proteins, TANGO 331 proteins, and TANGO 332 proteins). The isolated nucleic acid molecules of the

25   invention can be used to express proteins (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (e.g., in a biological sample) or a genetic lesion, and to modulate activity of a polypeptide of the invention. In addition, the polypeptides of the invention can be used to screen drugs or compounds which modulate activity or expression of a polypeptide of the invention as well as to treat

disorders characterized by insufficient or excessive production of a protein of the invention or production of a form of a protein of the invention which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies of the invention can be used to detect and isolate a protein of the and modulate activity of a protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A.   Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to polypeptide of the invention or have a stimulatory or inhibitory effect on, for example, expression or activity of a polypeptide of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a polypeptide of the invention or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al.

(1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds can be presented in solution (e.g., Houghten (1992) *Bio/Techniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent NOS. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382; and Felici (1991) *J. Mol. Biol.* 222:301-310).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radio-emission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the

polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

5   In another embodiment, the assay involves assessment of an activity characteristic of the polypeptide, wherein binding of the test compound with the polypeptide or a biologically active portion thereof alters (i.e., increases or decreases) the activity of the polypeptide.

          In another embodiment, an assay is a cell-based assay comprising
10   contacting a cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide or a
15   biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide to bind to or interact with a target molecule or to transport molecules across the cytoplasmic membrane.

          Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described
20   above for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide (e.g., a polypeptide of the invention binds or interacts with in nature, for example, a molecule on the surface of a cell which expresses the selected protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell
25   membrane or a cytoplasmic molecule. A target molecule can be a polypeptide of the invention or some other polypeptide or protein. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (e.g., a signal generated by binding of a compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular

protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., an mRNA, intracellular Ca^{2+}, diacylglycerol, IP3, and the like), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (e.g., a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the

polypeptide can be accomplished, for example, by determining the ability of the polypeptide to bind to a target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to further modulate the target molecule. For example, the catalytic activity, the enzymatic activity, or both, of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting a polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the polypeptide to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it can be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

In one or more embodiments of the above assay methods of the present invention, it can be desirable to immobilize either the polypeptide of the invention or its target molecule to facilitate separation of complexed from non-complexed forms of

one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione Sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or A polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptide of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes,

in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target

5   molecule.

In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a candidate compound and the expression of the selected mRNA or protein (i.e., the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is

10   determined. The level of expression of the selected mRNA or protein in the presence of the candidate compound is compared to the level of expression of the selected mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression of the polypeptide of the invention based on this comparison. For example, when expression of the selected mRNA or

15   protein is greater (i.e., statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of the selected mRNA or protein expression. Alternatively, when expression of the selected mRNA or protein is less (i.e., statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an

20   inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods described herein.

In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.,* U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993)

25   *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Bio/Techniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with the polypeptide of the invention and modulate activity of the polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide

- 158 -

of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

5

B.   Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

15   1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, nucleic acid molecules described herein or fragments thereof, can be used to map the location of the corresponding genes on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 base pairs in length) from the sequence of a gene of the invention. Computer analysis of the sequence of a gene of the invention can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will

yield an amplified fragment. For a review of this technique, see D'Eustachio et al. ((1983) *Science* 220:919-924).

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be

5   assigned per day using a single thermal cycler. Using the nucleic acid sequences of the invention to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a gene to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening

10   with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries. Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. For a review of this technique, see Verma et al. (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New

15   York, 1988)).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to non-coding regions of the genes actually are preferred for mapping purposes.

20   Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance

25   in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a gene of the invention can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Furthermore, the nucleic acid sequences disclosed herein can be used to perform searches against "mapping databases", e.g., BLAST-type search, such that the chromosome position of the gene is identified by sequence homology or identity with known sequence fragments which have been mapped to chromosomes.

A polypeptide and fragments and sequences thereof and antibodies which bind specifically with such polypeptides/fragments can be used to map the location of the gene encoding the polypeptide on a chromosome. This mapping can be performed by specifically detecting the presence of the polypeptide/fragments in members of a panel of somatic cell hybrids between cells obtained from a first species of animal from which the protein originates and cells obtained from a second species of animal, determining which somatic cell hybrid(s) expresses the polypeptide, and noting the chromosome(s) of the first species of animal that it contains. For examples of this technique (see Pajunen et al., 1988, Cytogenet. Cell Genet. 47:37-41 and Van Keuren et al., 1986, Hum. Genet. 74:34-40). Alternatively, the presence of the polypeptide in the somatic cell hybrids can be determined by assaying an activity or property of the polypeptide (e.g., enzymatic activity, as described in Bordelon-Riser et al., 1979, Som. Cell Genet. 5:597-613 and Owerbach et al., 1978, Proc. Natl. Acad. Sci. USA 75:5640-5644).

2. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the non-coding regions, fewer sequences are necessary to differentiate individuals. The

- 162 -

non-coding sequences of SEQ ID NO: 1, 9, 33, 38, 46, 54, 59, and 81 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a non-coding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO: 2, 10, 34, 39, 47, 55, 60, 82, and 92 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from the nucleic acid sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Gene Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e., another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to non-coding regions are particularly appropriate for this use as greater numbers of polymorphisms occur in

the non-coding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the nucleic acid sequences of the invention or portions thereof, e.g., fragments derived from non-coding regions having a length of at least 20 or 30 bases.

5   The nucleic acid sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by

10   species and/or by organ type.

C.   Predictive Medicine

       The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring

15   clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining expression of a polypeptide or nucleic acid of the invention and/or activity of a polypeptide of the invention, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted

20   with a disease or disorder, or is at risk of developing a disorder, associated with aberrant expression or activity of a polypeptide of the invention. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, mutations in a gene of the invention can be

25   assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with aberrant expression or activity of a polypeptide of the invention.

As an alternative to making determinations based on the absolute expression level of a selected gene, determinations can be based on normalized expression levels of the gene. A gene expression level is normalized by correcting the absolute expression level of the gene (e.g., an INTERCEPT 217, INTERCEPT 297,

5   TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332 gene as described herein) by comparing its expression to expression of a gene for which expression is not believed to be co-regulated with the gene of interest, e.g., a housekeeping gene that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene. Such normalization allows

10   comparison of the expression level in one sample, e.g., a patient sample, with the expression level in another sample, e.g., a sample obtained from a patient known not to be afflicted with a disease or condition, or between samples obtained from different sources.

Alternatively, the expression level can be assessed as a relative expression level. To assess a relative expression level for a gene (e.g., an INTERCEPT

15   217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332 gene, as described herein), the level of expression of the gene is determined for 10 or more samples (preferably 50 or more samples) of different isolates of cells in which the gene is believed to be expressed, prior to assessing the

20   level of expression of the gene in the sample of interest. The mean expression level of the gene detected in the large number of samples is determined, and this value is used as a baseline expression level for the gene. The expression level of the gene assessed in the test sample (i.e., its absolute level of expression) is divided by the mean expression value to yield a relative expression level. Such a method can identify

25   tissues or individuals which are afflicted with a disorder associated with aberrant expression of a gene of the invention.

Preferably, the samples used in the baseline determination are generated either using cells obtained from a tissue or individual known to be afflicted with a disorder (e.g., a disorder associated with aberrant expression of one of the

- 165 -

INTERCEPT 217, INTERCEPT 297, TANGO 276, TANGO 292, TANGO 325, TANGO 331, or TANGO 332 genes) or using cells obtained from a tissue or individual known not to be afflicted with the disorder. Alternatively, levels of expression of these genes in tissues or individuals known to be or not to be afflicted with the disorder can

5   be used to assess whether the aberrant expression of the gene is associated with the disorder (e.g., with onset of the disorder, or as a symptom of the disorder over time).

Another aspect of the invention provides methods for expression of a nucleic acid or polypeptide of the invention or activity of a polypeptide of the invention in an individual to thereby select appropriate therapeutic or prophylactic agents for that

10   individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent).

Yet another aspect of the invention pertains to monitoring the influence

15   of agents (e.g., drugs or other compounds) on the expression or activity of a polypeptide of the invention in clinical trials. These and other agents are described in further detail in the following sections.

1.   Diagnostic Assays

20   An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention such that the presence of a polypeptide or nucleic acid

25   of the invention is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA encoding a polypeptide of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a full-length cDNA, such as the nucleic acid of SEQ ID NO: 1, 9, 33, 38, 46, 54, 59, 62, or 81, or a portion

thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a polypeptide of the invention. Other suitable probes for use in the diagnostic assays of the invention are described herein.

5   A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide of the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of a polypeptide of the invention include introducing into a subject a labeled antibody directed against the polypeptide. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA

molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a polypeptide of the invention or mRNA or genomic DNA encoding a polypeptide of the invention, such that the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide is detected in the biological sample, and comparing the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the control sample with the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of a polypeptide of the invention (e.g., one of the disorders described in the section of this disclosure wherein the individual polypeptide of the invention is discussed). For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide or mRNA encoding the polypeptide in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (e.g., an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

- 168 -

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule encoding a polypeptide of the invention.

5   The kit can also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit can also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual

10   container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide.

2.   Prognostic Assays

15   The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of

20   developing a disorder associated with aberrant expression or activity of a polypeptide of the invention (e.g., one of the disorders described in the section of this disclosure wherein the individual polypeptide of the invention is discussed). Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing such a disease or disorder. Thus, the present invention provides a method in which a

25   test sample is obtained from a subject and a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the polypeptide. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest.

- 169 -

For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (e.g., agents of a type which decrease activity of the polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a polypeptide of the invention in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (e.g., wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized aberrant expression or activity of a polypeptide of the invention. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding the polypeptide of the invention, or the mis-expression of the gene encoding the polypeptide of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from the gene; 2) an addition of one or more nucleotides to the gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern

- 170 -

of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; 8) a non-wild type level of the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post-translational modification of the protein encoded by the gene. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a gene.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting·point mutations in a gene (*see, e.g.*, Abravaya et al. (1995) *Nucleic Acids Res*. 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. PCR and/or LCR can be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self-sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a selected gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, (optionally) amplified, digested with one or more restriction endonucleases, and fragment length sizes are determined by gel

5   electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

10   In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al. (1996) *Human Mutation* 7:244-255; Kozal et al. (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two-dimensional arrays containing

15   light-generated DNA probes as described in Cronin et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the

20   characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions

25   known in the art can be used to directly sequence the selected gene and detect mutations by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also

contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Bio/Techniques* 19:448), including sequencing by mass spectrometry (*see, e.g.*, PCT Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl.*

5   *Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in a selected gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the technique of mismatch cleavage entails providing heteroduplexes

10   formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to base pair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions,

15   and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions.

In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of

20   mutation. *See, e.g.*, Cotton et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so

25   called DNA mismatch repair enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a selected

sequence, e.g., a wild-type sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

5   In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2766; *see also* Cotton (1993) *Mutat. Res.* 285:125-144; Hayashi (1992) *Genet.*
10   *Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to re-nature. The secondary structure of single-stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments can be labeled or detected with labeled probes. The sensitivity of the
15   assay can be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet.* 7:5).

20   In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a 'GC clamp' of approximately 40
25   base pairs of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers can be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification can be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification can carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatching can prevent or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition, it can be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). Amplification can also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein can be performed, for example, using pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which can be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue,

preferably peripheral blood leukocytes, in which the polypeptide of the invention is expressed can be utilized in the prognostic assays described herein.

3. Pharmacogenomics

Agents, or modulators which have a stimulatory or inhibitory effect on activity or expression of a polypeptide of the invention as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of a polypeptide of the invention, expression of a nucleic acid of the invention, or mutation content of a gene of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. *See, e.g.*, Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in

- 176 -

which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of a polypeptide of the invention, expression of a nucleic acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or

expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drug compounds) on the expression or activity of a polypeptide of the invention (e.g., the ability to modulate aberrant cell proliferation chemotaxis, and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase gene expression, protein levels, or protein activity, can be monitored in clinical trials of subjects exhibiting decreased gene expression, protein levels, or protein activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or protein activity. In such clinical trials, expression or activity of a polypeptide of the invention and preferably, that of other polypeptide that have been implicated in for example, a cellular proliferation disorder, can be used as a marker of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including those of the invention, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates activity or expression of a polypeptide of the invention (e.g., as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of a gene of the invention and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of a gene of the invention or other genes. In this way, the gene

expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state can be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or nucleic acid of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level the of the polypeptide or nucleic acid of the invention in the post-administration samples; (v) comparing the level of the polypeptide or nucleic acid of the invention in the pre-administration sample with the level of the polypeptide or nucleic acid of the invention in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent can be desirable to increase the expression or activity of the polypeptide to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent can be desirable to decrease expression or activity of the polypeptide to lower levels than detected, i.e., to decrease the effectiveness of the agent.

D.   Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of a polypeptide of the invention and/or in which the polypeptide of the invention is involved. Disorders characterized by aberrant expression or activity of the polypeptides of the invention are described elsewhere in this disclosure.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant expression or activity of a polypeptide of the invention, by administering to the subject an agent which modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease which is caused or contributed to by aberrant expression or activity of a polypeptide of the invention can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrance, for example, an agonist or antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of a polypeptide of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. Examples of such stimulatory agents include the active polypeptide of the invention and a nucleic acid molecule encoding the polypeptide of the invention that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of the polypeptide of the invention. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by

- 180 -

administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a polypeptide of the invention. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) expression or activity. In another embodiment, the method involves administering a polypeptide of the invention or a nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the polypeptide.

Stimulation of activity is desirable in situations in which activity or expression is abnormally low or down-regulated and/or in which increased activity is likely to have a beneficial effect, e.g., in wound healing. Conversely, inhibition of activity is desirable in situations in which activity or expression is abnormally high or up-regulated and/or in which decreased activity is likely to have a beneficial effect.

The contents of all references, patents, and published patent applications cited throughout this application are hereby incorporated by reference.

Deposits of Clones

Clones encoding the proteins of the invention were deposited with the American Type Culture Collection (ATCC®, 10801 University Boulevard, Manassas, VA 20110-2209) on April 27, 1999 and May 27, 1999. These deposits will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. Each of these deposits was made merely as a convenience to those of skill in the art. These deposits are not an admission that a deposit is required under 35 U.S.C. §112.

Clones comprising cDNA molecules encoding human INTERCEPT 217, human INTERCEPT 297, human TANGO 325, and human TANGO 331 were deposited with ATCC® on May 28, 1999, as part of a composite deposit representing a

mixture of five strains, each carrying one recombinant plasmid harboring a particular cDNA clone. This deposit was assigned Accession Number PTA-147

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture is streaked out to single colonies on nutrient medium (e.g., Luria broth plates) supplemented with 100 micrograms per milliliter ampicillin, single colonies grown, and then plasmid DNA is extracted using a standard mini-preparation procedure. Next, a sample of the DNA mini-preparation is digested using a combination of the restriction enzymes *SalI*, *NotI*, and *SmaI*, and the resultant products are resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

1. human INTERCEPT 217 (clone EpT217): 2.9 kilobases

2. human INTERCEPT 297 (clone EpT297): 1.2 kilobases and 0.3 kilobases (human INTERCEPT 297 has a *SmaI* cut site at about base pair 1183).

3. human TANGO 325 (clone EpT325): 2.2 kilobases

4. human TANGO 331 (clone EpT331): 1.4 kilobases

The identity of the strains can be inferred from the fragments liberated.

Human TANGO 276, human TANGO 292, and human TANGO 332 were each deposited as single deposits. Their clone names, deposit dates, and accession numbers are as follows:

1. human TANGO 276: clone EpT276 was deposited with ATCC® on May 28, 1999, and was assigned Accession Number PTA-150.

2. human TANGO 292: clone EpT292 was deposited with ATCC® on April 28, 1999, and was assigned Accession Number 207230.

3. human TANGO 332: clone EpT332 was deposited with ATCC® on May 28, 1999, and was assigned Accession Number PTA-151.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.